



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : A01N 37/18, 43/04, C12Q 1/00, 1/02, 1/68, C12N 5/00, 5/06, 15/00, 15/06, 15/09, 15/10, 15/11, G01N 33/53	A2	(11) International Publication Number: WO 98/54963 (43) International Publication Date: 10 December 1998 (10.12.98)
(21) International Application Number: PCT/US98/11422 (22) International Filing Date: 4 June 1998 (04.06.98) (30) Priority Data: 60/048,915      6 June 1997 (06.06.97)      US 60/048,882      6 June 1997 (06.06.97)      US (Continued on the following page) (71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): YOUNG, Paul [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). GREENE, John, M. [US/US]; 872 Diamond Drive, Gaithersburg, MD 20878 (US). FERRIE, Ann, M. [US/US]; 13203 L Astoria Hill Court, Germantown, MD 20874 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). HU, Jing-Shan [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316,		Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; 14920 Mt. Nebo Road, Poolesville, MD 20837 (US). MOORE, Paul, A. [GB/US]; Apartment 104, 1908 Holly Ridge Drive, McLean, VA 22102 (US). SHI, Yanggu [CN/US]; 437 West Side Drive, Gaithersburg, MD 20878 (US). FLORENCE, Charles [US/US]; (US). FLORENCE, Kimberly [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). LAFLEUR, David, W. [US/US]; 1615 Q Street, N.W. #807, Washington, DC 20009 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FAN, Ping [CN/US]; Apartment 302, 335 West Side Drive, Gaithersburg, MD 20878 (US). WEI, Ying-Fei [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). FISCHER, Carrie, L. [US/US]; 5810 Hall Street, Burke, VA 22015 (US). SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). LI, Yi [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). ZENG, Zhizhen [CN/US]; 13950 Saddleview Drive, Gaithersburg, MD 20878 (US). KYAW, Hla [MM/US]; 520 Sugarbush Circle, Frederick, MD 21703 (US). YU, Guo-Liang [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US). CARTER, Kenneth, C. [US/US]; 11601 Brandy Hall Lane, North Potomac, MD 20878 (US). (74) Agents: HOOVER, Kenley, K. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published With declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority.
(54) Title: 207 HUMAN SECRETED PROTEINS		

(Continued)

60/048.892	6 June 1997 (06.06.97)	US	60/057.651	5 September 1997 (05.09.97)	US
60/048.901	6 June 1997 (06.06.97)	US	60/057.769	5 September 1997 (05.09.97)	US
60/048.900	6 June 1997 (06.06.97)	US	60/057.643	5 September 1997 (05.09.97)	US
60/048.893	6 June 1997 (06.06.97)	US	60/057.645	5 September 1997 (05.09.97)	US
60/048.964	6 June 1997 (06.06.97)	US	60/057.668	5 September 1997 (05.09.97)	US
60/048.884	6 June 1997 (06.06.97)	US	60/057.635	5 September 1997 (05.09.97)	US
60/048.894	6 June 1997 (06.06.97)	US	60/057.627	5 September 1997 (05.09.97)	US
60/048.971	6 June 1997 (06.06.97)	US	60/057.667	5 September 1997 (05.09.97)	US
60/048.885	6 June 1997 (06.06.97)	US	60/057.666	5 September 1997 (05.09.97)	US
60/049.375	6 June 1997 (06.06.97)	US	60/057.764	5 September 1997 (05.09.97)	US
60/048.881	6 June 1997 (06.06.97)	US	60/057.644	5 September 1997 (05.09.97)	US
60/048.880	6 June 1997 (06.06.97)	US	60/057.765	5 September 1997 (05.09.97)	US
60/048.896	6 June 1997 (06.06.97)	US	60/057.762	5 September 1997 (05.09.97)	US
60/049.020	6 June 1997 (06.06.97)	US	60/057.775	5 September 1997 (05.09.97)	US
60/048.876	6 June 1997 (06.06.97)	US	60/057.634	5 September 1997 (05.09.97)	US
60/048.895	6 June 1997 (06.06.97)	US	60/057.777	5 September 1997 (05.09.97)	US
60/049.019	6 June 1997 (06.06.97)	US	60/057.628	5 September 1997 (05.09.97)	US
60/048.916	6 June 1997 (06.06.97)	US	60/057.776	5 September 1997 (05.09.97)	US
60/048.970	6 June 1997 (06.06.97)	US	60/057.760	5 September 1997 (05.09.97)	US
60/048.972	6 June 1997 (06.06.97)	US	60/057.761	5 September 1997 (05.09.97)	US
60/048.949	6 June 1997 (06.06.97)	US	60/057.771	5 September 1997 (05.09.97)	US
60/048.974	6 June 1997 (06.06.97)	US	60/057.770	5 September 1997 (05.09.97)	US
60/048.883	6 June 1997 (06.06.97)	US	60/057.649	5 September 1997 (05.09.97)	US
60/048.897	6 June 1997 (06.06.97)	US	60/057.774	5 September 1997 (05.09.97)	US
60/048.898	6 June 1997 (06.06.97)	US	60/057.648	5 September 1997 (05.09.97)	US
60/049.373	6 June 1997 (06.06.97)	US	60/057.642	5 September 1997 (05.09.97)	US
60/048.917	6 June 1997 (06.06.97)	US	60/057.629	5 September 1997 (05.09.97)	US
60/048.962	6 June 1997 (06.06.97)	US	60/057.778	5 September 1997 (05.09.97)	US
60/048.878	6 June 1997 (06.06.97)	US	60/057.763	5 September 1997 (05.09.97)	US
60/049.374	6 June 1997 (06.06.97)	US	60/057.584	5 September 1997 (05.09.97)	US
60/048.875	6 June 1997 (06.06.97)	US	60/057.654	5 September 1997 (05.09.97)	US
60/048.899	6 June 1997 (06.06.97)	US	60/057.646	5 September 1997 (05.09.97)	US
60/048.877	6 June 1997 (06.06.97)	US	60/057.662	5 September 1997 (05.09.97)	US
60/048.963	6 June 1997 (06.06.97)	US	60/057.650	5 September 1997 (05.09.97)	US
			60/057.661	5 September 1997 (05.09.97)	US
			60/057.647	5 September 1997 (05.09.97)	US
			60/070.923	18 December 1997 (18.12.97)	US

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						



## 207 Human Secreted Proteins

### *Field of the Invention*

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

### *Background of the Invention*

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

### *Summary of the Invention*

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

### *Detailed Description*

#### Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences; the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard,  
5 Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained  
10 in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the  
15 filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages  
20 of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH<sub>2</sub>PO<sub>4</sub>; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even  
25 lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include  
30 Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such  
35 as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

5       The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and  
10   double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability  
15   or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

      The polypeptide of the present invention can be composed of amino acids joined  
20   to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs,  
25   as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be  
30   branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a  
35   nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS -  
STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W.  
H. Freeman and Company, New York (1993); POSTTRANSLATIONAL  
COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic  
Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990);  
Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

## Polynucleotides and Polypeptides of the Invention

### FEATURES OF PROTEIN ENCODED BY GENE NO: 1

This gene is expressed primarily in melanocytes and, to a lesser extent, in testes, ovary, kidney and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer, disorders of neural crest derived cells including pigmentation defects, melanoma, reproductive organ defects, and defects of the kidney. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin,

reproductive, and renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating disorders that arise from alterations in the number or fate of neural crest derived cells including cancers such as melanoma and defects of the developing reproductive system.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 2

This gene is expressed primarily in infant brain and fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental disorders of the brain or lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating or diagnosing disorders associated with abnormal proliferation of cells in the Central nervous system and developing lung.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in breast lymph node and to a lesser extent in ovarian cancer and chondrosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune responses such as inflammation or immune surveillance for

tumors. This gene may be important for inflammatory responses associated with tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 236 as residues: Lys-45 to Val-50, Lys-69 to Arg-76.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of immune responses including those associated with tumor-induced inflammation.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 4**

This gene is expressed primarily in T-cells and T-cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological diseases involving T-cells such as inflammation, autoimmunity, and cancers including T-cell lymphomas. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of T-cells and other cells of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating T-cell based disorders such as inflammatory diseases, autoimmune disease and tumors including T-cell lymphomas.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 5**

This gene is expressed primarily in activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation, autoimmunity, infection, or disorders involving activation of monocytes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 238 as residues: Asp-19 to Arg-31.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing or treating diseases that result in activation of monocytes including infections, inflammatory responses or autoimmune diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 6**

The translation product of this gene shares sequence homology with terminal deoxynucleotidyltransferase which is thought to be important in catalyzing the elongation of oligo- or polydeoxynucleotide chains.

This gene is expressed primarily in activated human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, particularly those of the blood such as leukemia and deficiencies in neutrophils such as neutropenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having



such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to terminal deoxynucleotidyltransferase indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and differential diagnosis of acute leukemia's. Alternatively, this gene may function in the proliferation of neutrophils and be useful as a treatment for neutropenia, for example, following neutropenia as a result of chemotherapy.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 7

The contig exhibits a reasonable homology to the human chorionic gonadotropic (HCG) analogue-GT beta-subunit as disclosed in U.S. Patent No. 5,508,261 and PCT Publication No. WO 92/22568. There is a high degree of conservation of the structurally important cysteine residues in these identities.

This gene is expressed primarily in IL-1 and LPS induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 8

This gene is expressed primarily in IL-1- and LPS-induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 241 as residues: Ser-14 to Pro-22, Leu-43 to Val-53.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 9

This gene is expressed primarily in IL-1 and LPS induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 242 as residues: Tyr-22 to His-35.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth

factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 10

5 This gene is expressed primarily in activated T-cells and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune dysfunctions including cancer of the T lymphocytes and autoimmune disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of immune disorders particularly of T-cell origin and may act as a growth factor for particular subsets of T-cells such as CD4 positive cells which would make this a useful therapeutic for the treatment of HIV and other immune compromising illnesses.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 11

This gene is expressed primarily in fetal tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of many developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor or differentiation factor for particular cell types in the developing fetus and may be useful in replacement or other types of therapy in cases where the gene is expressed aberrantly.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 12

This gene is expressed primarily in T-cells and to a lesser extent in tumor tissue including glioblastoma, meningioma, and Wilm's tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system including autoimmune conditions such as rheumatoid arthritis, inflammatory disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 245 as residues: Thr-9 to Ser-14.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis/ modulation of immune function disorders, including rheumatoid arthritis and inflammatory responses.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene is expressed primarily in placenta and to a lesser extent in fetal liver and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hematological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of

disorders of the above tissues or cells, particularly of the hematological and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample  
5 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or  
10 progenitor cells in the treatment of chemotherapy patients or kidney disease.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 14

This gene is expressed primarily in stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as  
15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hematopoietic disorders including cancer, neutropenia, anemia, and thrombocytopenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of  
20 the above tissues or cells, particularly of the hematopoietic and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,  
25 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or progenitor cells, in particular following chemotherapy treatment.

30

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 15

The translation product of this gene shares sequence homology with epsilon-COP from *Bos taurus* which is thought to be important as a component of coatamer, a complex of seven proteins, that is the major component of the non-clathrin membrane  
35 coat. Preferred polypeptides encoded by this gene comprise the following amino acid sequences:

MAPPAPGPASGGSGEVDELFDVKNAFYIGSYQQCINEAXXVKLSSPERDVERD

VFLYRAYLAQRKFGVVLDEIKPSSAPELQAVRMFADYLAHESRRDSIVAELDRE  
MSRSXDVTNTTFLMLAASIYLHDQNPDAALRALHQGDSLECTAMTVQILLKLD  
RLDLARKELKRMQDLDEDATLTQLATAWVSLATGGEKLQDAYYTFQEMADKCS  
PTLLLLNGQAACHMAQGRWEAAEGLLQEALDKDSGYPETLVNLIVLSQHLGKP  
5 PEVTNRYLSQLKDAHRSHPIKEYQAKENDFDRLVLQYAPSAEAGPELSGP  
(SEQ ID NO:458); or RDVERDVFLYRAYLAQRKFGVVLDEIKPSSAPELQAVRMF  
ADYLAHESRRDSIVAELDREMSRSXDVTNTTFLMLAASIYLHDQNPDAALRALH  
QGDSLECTAMTVQILLKLDRLDLARKELKRMQDLDEDATLTQLATAWVSLATG  
GEKLQDAYYTFQEMADKCSPTLLLLNGQAACHMAQGRWEAAEGLLQEALDKD  
10 SGYPETLVNLIVLSQHLGKPPPEVTNRYLSQLKDAHRSHPIKEYQAKENDFDRL  
VLQYAPSA (SEQ ID NO:459).

This gene is expressed primarily in activated monocytes and T-cells, and to a lesser extent in multiple other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as  
15 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
not limited to, immunomodulation, specifically relating to transport problems in these  
cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
providing immunological probes for differential identification of the tissue(s) or cell  
20 type(s). For a number of disorders of the above tissues or cells, particularly of the  
immune, expression of this gene at significantly higher or lower levels may be routinely  
detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,  
serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample  
taken from an individual having such a disorder, relative to the standard gene  
25 expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
individual not having the disorder.

The tissue distribution and homology to epsilon-COP indicates that  
polynucleotides and polypeptides corresponding to this gene are useful for treating  
/diagnosing problems with the cellular transport of proteins that may result in  
30 immunologic dysfunction.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 16

The translation product of this gene shares sequence homology with an RNA  
helicase which is thought to be important in polynucleotide metabolism. The translation  
35 product of this contig exhibits good homology to the LbeIF4A antigen of *Leishmania*  
*braziliensis*. The LbeIF4A antigen, or immunogenic portions of it, can be used to  
induce protective immunity against leishmaniasis, specifically *L. donovani*, *L. chagasi*.

L. infantum, L. major, L. braziliensis, L. panamensis, L. tropica and L. guyanensis. It can also be used diagnostically to detect Leishmania infection or to stimulate a cellular and/or humoral immune response or to stimulate the production of interleukin-12.

This gene is expressed primarily in colon cancer and to a lesser extent in pituitary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of cancers particularly of the colon. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gastrointestinal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 249 as residues: Glu-93 to Ala-98, Gln-150 to Leu-156, Leu-220 to Leu-231, Leu-268 to Arg-273, Val-324 to Pro-341, Arg-372 to Asn-380, Ser-405 to Gly-410, Phe-426 to Ala-433, Glu-458 to Asp-470, Arg-506 to Ser-547.

The tissue distribution and homology to RNA helicase indicates that polynucleotides and polypeptides corresponding to this gene are useful for development of diagnostic tests for colon cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 17

The translation product of this contig has sequence homology to a cytoplasmic protein that binds specifically to JNK designated the JNK interacting protein-1 or JIP-1 in mice. JIP-1 caused cytoplasmic retention of JNK and inhibition of JNK-regulated gene expression.

This gene is expressed primarily in brain including pituitary cerebellum frontal cortex, fetal brain and to a lesser extent in the kidney cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of the central nervous system disorders including ischemia, epilepsy, Parkinson's disease, and schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

probes for differential identification of the tissue(s)-or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Furthermore, the translation product of this contig may suppress the effects of the JNK signaling pathway on cellular proliferation, including transformation by the Bcr-Abl oncogene. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 250 as residues: Pro-6 to Ser-26, Ala-30 to Asp-41, Gly-55 to Ser-61, Gly-74 to Thr-80, Tyr-117 to Ala-123, Tyr-167 to Asp-172, Ala-212 to Cys-223, Pro-239 to Tyr-244.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for enhanced survival and/or differentiation of neurons as a treatment for neurodegenerative disease.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 18

The translation product of this gene shares sequence homology with a liver stage antigen from a protozoan parasite.

This gene is expressed primarily in fetal tissue and to a lesser extent in activated T-cells and other immune cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities and diseases of immune function. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.



The tissue distribution and homology to a protozoan antigen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/immune modulation of parasitic infections.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 19

Preferred polypeptide encoded by this gene comprise the following polypeptide sequences:

MKAIGIEPSLATYHHIIRLFDQPGDPLKRSSFIIYDIMNELMGKRFSPKD  
 PDDDKFFQSAMSICSSLRDLELAYQVHGLLKTGDNWKFIPDQHRNFYYSKFF  
 10 DLICLMEQIDVTLKWYEDLIPSA YFPHSQTMIHLLQALDVANRLEVIPKTWER  
 (SEQ ID NO:460); and/or KDSKEYGHTFRSDLREEILMLMARDKHPPELQVAF  
 ADCAADIKSAYESQPIRQTAQDWPATSLNCIAILFLRAGRTQEAWKMLGLFRKH  
 NKIPRSELLNELMDSAKVSNPSQAIEVVELASAFSLPICEGLTQRMVMSDFAINQ  
 EQKEALSNTALTSDSDTDSSSDSDSDTSEGK (SEQ ID NO:461). Polynucleotides

15 encoding such polypeptides are also provided.

This gene is expressed primarily in stromal and CD34 depleted bone marrow cells and to a lesser extent in tissues of embryonic origin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
 20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of hematologic origin including cancers and immune dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of  
 25 the hematopoietic and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily  
 30 fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 252 as residues: Ser-28 to Gln-34.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or progenitor cells which may be useful in the treatment of chemotherapy patients  
 35 suffering from neutropenia.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 20

Preferred polypeptide fragments can be found in an alternative open reading frame. These preferred polypeptides comprise the amino acid sequence:

MSSDNESDIEDLKLRLRDLKHLKEIQDLQSRQKHEIESLYTKLGKVPPAVI  
 5 IPPAAPLSGRRRRPTKSKGSKSSRSSLGNKSPQLSGNLSGQSAASVLHPQOTL  
 HPPGNIPESGQNQLLQPLKPS SSDNLYSAFTSDGAISVPSLSAPGQGTSTNTV  
 GATVNSQAAQAQPPAMTSSRKGTFDDHLKLV DNWARDAMNLSGRRGSKGH  
 MNYEGPGMARKFSAPGQLCISMTSNLGGSA PISAASATSLGHFTKSMCPPQQY  
 GFPATPFGAQWSGTGGPAPQLGQFQPVGTASLQNFNISNLQKSISNPPGSNL  
 10 RTT (SEQ ID NO:462); IQDLQSRQKHEIESLYTKLGKVPPAVIIPPAAPLSGRRRR  
 PTKSKGSKSSRSSLGNKSPQLSGNLSGQSAASVLHPQOTLHPPGNIPESGQN  
 QLLQPLKPS SSDNLYSAFTSDGAISVPSLSAPGQGTSTT (SEQ ID NO:463);  
 TSDGAISVPSLSAPGQGTSTNTVGATVNSQAAQAQPPAMTSSRKGTFDDHL  
 (SEQ ID NO:464); KGHMNYEGPGMARKFSAPGQLCISMTSNLGGSA PISAAS  
 15 ATSLGHFTK (SEQ ID NO:465); QPLKPS SSDNLYSAFTSDGAISVPSLSAPG  
 (SEQ ID NO:466). Also preferred are polynucleotide fragments encoding these  
 polypeptide fragments.

This gene is expressed in fetal liver and tissues associated with the CNS.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 20 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions, which include, but are  
 not limited to, liver and CNS diseases. Similarly, polypeptides and antibodies directed  
 to these polypeptides are useful in providing immunological probes for differential  
 identification of the tissue(s) or cell type(s). For a number of disorders of the above  
 25 tissues or cells, particularly of the liver and CNS, expression of this gene at  
 significantly higher or lower levels may be routinely detected in certain tissues (e.g.,  
 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial  
 fluid or spinal fluid) or another tissue or cell sample taken from an individual having  
 such a disorder, relative to the standard gene expression level, i.e., the expression level  
 30 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred  
 epitopes include those comprising a sequence shown in SEQ ID NO: 253 as residues:  
 Gln-26 to Lys-34.

The tissue distribution indicates that polynucleotides and polypeptides  
 corresponding to this gene are useful for diagnosis and treatment for liver diseases such  
 35 as hepatocellular carcinomas and diseases of the CNS.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 21**

In an alternative reading frame, this gene shows sequence homology to two recently cloned genes, karyopherin beta 3 and Ran\_GTP binding protein 5. (See Accession Nos: gil2102696 and gnlIPIDe328731.) The Ran\_GTP binding protein is related to importin-beta, the key mediator of nuclear localization signal (NLS)-dependent nuclear transport. Based on homology, it is likely that this gene may activity similar to the RAN\_GTP binding protein. Preferred polypeptide fragments comprise the amino acid sequence: VRVAAAESMXLLLECAXVRGPEYLTQMWHFMCDALIKA IGTEPDSDVLSEIMHSFAK (SEQ ID NO:467). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in thymus tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for immune disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 22**

This gene is expressed primarily in prostate and osteoclastoma tissues. Preferred polypeptide fragments also comprise the amino acid sequence: MEINNQNCFIVIDLVRTVMENGVEGLLIFGAFLPESWLIGVRCSSPEPKALLLIL AHSQKRRLDGWSFIRHLRVHYCVSLTIHFS (SEQ ID NO:468). Also preferred are polynucleotide sequences encoding this polypeptide fragment.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bone and prostate diseases, and cancers, particularly of the bone and prostate. Similarly, polypeptides and antibodies directed to these polypeptides are

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone and prostate systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 255 as residues: Met-1 to Ser-11.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for bone and prostate disorders, especially cancers of those systems.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 23

This gene shares sequence homology with the FK506-binding protein (FKBP-13) family, a known cytosolic receptor for the immunosuppressants. Recently, another group has cloned a very similar gene, recognizing the homology to FK506-binding protein family, calling their gene FKBP23. (See Accession No. 2827255.)

This gene is expressed primarily in lymphoid tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample, especially for those susceptible to immune suppressant therapies and for diagnosis of diseases and conditions, which include, but are not limited to, immune suppressant disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 256 as residues: Ala-19 to Val-31, Arg-38 to Gly-49, Ala-61 to Lys-66, Tyr-68 to Pro-78, Gly-116 to Ala-121, Asp-154 to Ser-162, Glu-173 to Gln-186, Phe-194 to Gly-203. Pro-207 to Val-212.

The tissue distribution and homology to FKBP-12 and -13 indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for immune suppressant disorders.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 24

This gene is expressed primarily in the brain and in the retina. This gene maps to chromosome 8, and therefore can be used in linkage analysis as a marker for chromosome 8.

Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and ocular associated disease states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of  
15 disorders of the above tissues or cells, particularly of the disorders of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard  
20 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 257 as residues: Cys-34 to Asp-40.

The tissue distribution in retina indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and/or detection of eye disorders  
25 including blindness, color blindness, impaired vision, short and long sightedness, retinitis pigmentosa, retinitis proliferans, and retinoblastoma. Expression in the brain indicates a role in the is useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive  
30 disorder and panic disorder.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 25

This gene shows sequence homology to a newly identified class of proteins expressed in the nervous system, called stathmin family. (See Accession No. 2585991;  
35 see also Eur. J. Biochem. 248 (3), 794-806 (1997).) The stathmin family appears to be an ubiquitous phosphoprotein involved as a relay integrating various intracellular signaling pathways. These pathways affect cell proliferation and differentiation.

Preferred polypeptide fragments comprise the amino acid sequence:

QDKHAEVRKNKELKEEASR (SEQ ID NO:469); QQDLSPWAAPVGCPLXXASX  
TCHXLPLSGCLRRQSXSLPVVAXLCFWFSCPLASLFVPGQPCVTCPFPPLPFQD  
KHAEVRKNKELKEEASR (SEQ ID NO:470). Also preferred are the

5 polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

## 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 26

The polynucleotide sequence of this gene contains a domain similar to a Flt3 ligand peptide. Preferred polypeptide fragments comprise the amino acid sequence: PTRCCTTQPCRSSARRPCWVPMVPSPEGREXQPTCPS (SEQ ID NO:471). Thus, this gene may have activity as binding to Flt3 receptors, a process known to promote angiogenesis and/or lymphangiogenesis.

30 This gene is expressed in human tonsil, and to a lesser extent in teratocarcinoma, placenta, colon carcinoma, and fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the tonsil, as well as cancers, such as colon, reproductive, and kidney cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful.

in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tonsils, colon, reproductive organs, and kidneys, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,

5 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 259 as residues:

10 Pro-22 to Glu-33.

The tissue distribution in tonsil and several cancers and fetal tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the tonsil or colon, such as tonsillitis, inflammatory diseases involving nose and paranasal sinuses, especially during the

15 infection of influenza, adenoviruses, parainfluenza, rhinoviruses. The gene may also be useful in the diagnosis and treatment of neoplasms of nasopharynx or colon origins.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 27

In an alternative reading frame exists a large open reading frame that encodes a preferred polypeptide. Preferred polypeptide fragments comprise the amino acid

20 sequence:  
MKRSLNENSARSTAGCLPVPLFNQKKRNRQPLTSNPLKDDSGISTPSDNYDFP  
PLPTDWAWEA VNPEXAPVMKTVDTGQIPHSVSRPLRSQDSVFNSIQSNTGRSQ  
GGWSYRDGNKNTSLKTWXKNDFKPQCKRNTLVANDGKNSCPMSSGAQQQK  
25 QLRTPPEPPNLSRNKETELLRQTHSSKISGCTMRGLDKNSALQTLKPNFQQNQY  
KXQMLDDIPEDNTLKETSLYQLQFKEKASSLRISAVIESMKYWREHAQKTVLL  
FEVLAVLDSAVTPGPYYSKTFLMRDYGKNTLPCVFYEIDRELPRIRGRVHRCVG  
NYDQKKNIFQCVSVRPASVSEQKTFQAFVKIADVEMQYYINVMNET (SEQ ID  
NO:472); SQDSVFNSIQSNTGRSQGGWSYRDGNKNTSLKTWXKNDFKPQCKR  
30 (SEQ ID NO:473); NKETELLRQTHSSKISGCTMRGLDKNSALQTLKPNF (SEQ ID  
NO:474); SSLRISAVIESMKYWREHAQKTVLLFEVLAVLDSAVTPGPYYSKTFLM  
(SEQ ID NO:475); and PRLIRGRVHRCVGNYDQKKNIFQCVSVRPASVSEQK  
TFQAFV (SEQ ID NO:476).

This gene is expressed primarily in human testes.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, male reproductive disorders, including cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a hormone with reproductive or other systemic functions; contraceptive development; male infertility of testicular causes, such as Klinefelter's syndrome, varicocele, orchitis; male sexual dysfunctions; testicular neoplasms; and inflammatory disorders such as epididymitis.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 28

This gene is expressed primarily in apoptotic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases relating to T cells, as well as cancer in general. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the disorders of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune disorders. Moreover, since the gene was isolated from an apoptotic cell and based on the understanding of the relationship of apoptosis and cancer, it is likely that this gene may play a role in the genesis of cancer.



**FEATURES OF PROTEIN ENCODED BY GENE NO: 29**

This gene is expressed primarily in human tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 5 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions, which include, but are  
 not limited to, gastrointestinal disorders. Similarly, polypeptides and antibodies directed  
 to these polypeptides are useful in providing immunological probes for differential  
 identification of the tissue(s) or cell type(s). For a number of disorders of the above  
 10 tissues or cells, particularly of the gastrointestinal system, expression of this gene at  
 significantly higher or lower levels may be routinely detected in certain tissues (e.g.,  
 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial  
 fluid or spinal fluid) or another tissue or cell sample taken from an individual having  
 such a disorder, relative to the standard gene expression level, i.e., the expression level  
 15 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
 corresponding to this gene are useful for the diagnosis and treatment of gastrointestinal  
 diseases.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 30**

The translation product of this gene shares sequence homology with C44C1.2  
 gene product of *Caenorhabditis elegans* with unknown function. Preferred polypeptide  
 fragments comprise the amino acid sequence:

GVFRPCVCGRPASLTCSPLDPEVGPYCDTPTMRTLNFNLLWLALACSPVHTTLSK  
 25 SDAKKAASKTLLKSQFSDKPVQDRGLVVTDLKAESVVLEHRSYCSAKARDRH  
 FAGDVLGYVTPWNSHGVDVTKVFGSKFTQISPVWLQLKRRGREMF EVTGLHD  
 VDQGWMAVRKHAKGLHIVPRLLFEDWTYDDFRNVLDSEDEIEELSKTVVQVA  
 KNQHFDGFVVEVWNQLLSQKRVGLIHMLTHLAEALHQA RLLALLVIPPAITPGT  
 DQLGMFTHKEFEQLAPVLDGFSLMTYDYSTAHPGP NAPLSWVRACVQVLDP  
 30 KXKWRTKSSWGSTSMXWTRXPXDARXPVVGXR XIQXLKDHXP R MVLD SK  
 PQ (SEQ ID NO:477); TCSPLDPEVGPYCDTPTMRTLNFNLLWLALACSPVHTTLS  
 (SEQ ID NO:478); LVVTDLKAESVVLEHRSYCSAKARDRH FAGDVLGYVTPW  
 NSHGVDVTKVFGSKF (SEQ ID NO:479); REMFEVTGLHDVDQGWMAVRK  
 HAKGLHIVPRLLFEDWTYDDFRNVLDSEDE (SEQ ID NO:480); HFDGFVVEVW  
 35 NQLLSQKRVGLIHMLTHLAEALHQA RLLALLVIPPAITPGTDQLGM (SEQ ID  
 NO:481); DGFSLMTYDYSTAHPGP NAPLSWVRACVQVLDPKXKWRTKSSW  
 GST (SEQ ID NO:482). Also preferred are polynucleotide fragments encoding these

polypeptide fragments. This gene maps to human chromosome 11, and therefore is useful in linkage analysis as a marker for chromosome 11.

This gene is expressed primarily in human T cells and to a lesser extent in human colon carcinoma.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for  
10 differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from  
15 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 263 as residues: Leu-21 to Ala-30, Ser-38 to Asp-47, Pro-87 to Asp-94, Leu-197 to Thr-204, Pro-256 to Ser-262, Thr-277 to Arg-282, Thr-293 to Trp-303.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders and gastrointestinal diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 31

25 The translation product of this gene shares sequence homology with Ribosomal protein L11 of *Caenorhabditis elegans*. (See Accession No. 156201.) Preferred polypeptide fragments comprise the amino acid sequence:

ERGVSINQFCKEFNERTKDIKEGIPLPTKILVKPDRTFEIKIGQPTVSYFLKAAAG  
IEKGARQTGKEVAGLVTLKHVYELARIKAQDEAFALQDVPLSSVVRSIIGSARSL  
30 GIRVVKDLSSSEELAAFQKERAIFLAAQKEADLAAQEEAAKK (SEQ ID NO:483).

Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in human embryo tissue and to a lesser extent in human epithelioid sarcoma and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as  
35 reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development disorders and epithelial cell cancer. Similarly, polypeptides and antibodies

directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryonic and epithelial cell systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 264 as residues: Lys-34 to Gly-40.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of developmental disorders and epithelial cancer.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 32

This gene is expressed primarily in resting T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory and general immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders of immune system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 33

This gene is believed to reside on chromosome 1. Accordingly, polynucleotides derived from this gene are useful in linkage analysis as chromosome 1 markers.

This gene is expressed primarily in prostate and to a lesser extent in soares adult brain, human umbilical vein endothelial cells, and amniotic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate-related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the urinary system and nervous system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the diagnosis and treatment of disorders of the urinary and nervous systems.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 34

This gene shares sequence homology with R05G6.4 gene product. (See Accession No. gil1326338.) This gene also shares sequence homology with the cyclophilin-like protein CyP-60. (See Accession No. 1199598, see also Biochem. J. 314 (1), 313-319 (1996).) Preferred polypeptide fragments comprise the amino acid sequence:

AVYTYHEKKKDDTAASGYGTQNIRLSRDAVKDFDCCCLSLQPCHDPVVTDPGYL  
YEREAILEYILHQKKEIARQMKAIEKQRGTRREEQKELQRAASQDHVRGFLEKE  
SAIVSRP LNPFTAKALSGTSPDDVQPGPSVGPPSKDKDKVLPSFWIPSLTPEAK  
ATKLEKPSRTVTCMSGKPLRMSDLTPVHFTPLDSSVDRVGLITRERYVCAVT  
RDSLSNATPCAVLRPSGAVVTLECEKLRKDMVDPVTGDKLTDRDIIVLQRG  
(SEQ ID NO:484); YLYEREAILEYILHQKKEIARQMKAIEKQRGTRREEQKELQ  
RAASQDHVRGFLE (SEQ ID NO:485); and FTAKALSGTSPDDVQPGPSVGPP  
SKDKDKVLPSFWIPSLTPEAKATKLEKPSRTVTCMSGKPL (SEQ ID NO:486).

Also preferred are polynucleotide fragments that encode these polypeptide fragments.

This gene is expressed primarily in human testis and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive disorders and in particular testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system. Expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the male reproductive system and in particular of testicular cancer.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 35

The translation product of this gene shares sequence homology with Lpe5p of *Saccharomyces cerevisiae* which is thought to be important in the metabolism of phospholipids.

This gene is expressed primarily in liver and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and nervous systems expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 268 as residues: Pro-14 to Leu-20, Lys-28 to Asn-38, Arg-109 to Arg-114, Lys-119 to Asn-124, Glu-152 to Leu-157, Pro-172 to Val-180.

The tissue distribution and homology to Lpe5p of *Saccharomyces cerevisiae* indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of metabolic and nervous disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 36**

This gene shares sequence homology with the nuclear ribonucleoprotein U (HNRNP U), encoded by *C. elegans* (See Accession gill1703576.) Preferred polypeptide fragments comprise the amino acid sequence:

- 5 MDTSENRPENDVPEPPMPIADQVSNDDRPEGSVEDEEKKESSLPKSFKRKISVV  
SATKGVPAAGNSDTEGGQPGRKRRWGASTATTQKKPSISITTESLKSLIPDIKPL  
AGQEAVVDLHADDSETERNGDDGTHDKGLKICRTVTQVPAEGQENGQ  
REEEEEKEPEAEPPVPPQVSVEVALPPPAEHEVKKVTLGDTLTRRSISQQKSGV  
SITIDDPVRTAQVPSPPRGKISNIVHISNLVRPFTLGQLKELLGRTGTLVEEAFWI  
10 DKIKSHCFVTYSTVEEAVATRTALHGKWPQSNPKFLCADYAEQDEL DYHRGL  
LVDRPSETKTEEQGIPLHPPPPPPVQPPQHPRAEQREQERAVREQWAERERE  
MERRERTRSEREWDRDKVREGPRSRSRXRRRKERAKSKEKKSEKKEKAQE  
EPPAKLLDDLFRKTKAAPCIYWLPLTDSQIVQKEAERAERAKEREKRRKEQEEE  
EQKEREKEAERERNRQLEREKRREHSRERDRERERERERDRGDRDRDRERDRE  
15 RGRERDRRDTKRHSRSRSTPVRDRGGR (SEQ ID NO:488): Also preferred are  
the polynucleotide fragments encoding this polypeptide fragments.

This gene is expressed primarily in epididymus.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the male reproductive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of  
25 this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the  
30 disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of male reproductive disorders.

**35 FEATURES OF PROTEIN ENCODED BY GENE NO: 37**

This gene is expressed primarily in amygdala.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory diseases and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the amygdala, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of inflammatory diseases and reproductive disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 38

This gene shares sequence homology with human opsonin protein P35 fragment. (See Accession No. R94181.) The opsonin protein activates the phagocytosis of pathogenic microbes by phagocytic cells. Preferred polypeptide fragments comprise the amino acid sequence: GCDSCPPHLPREAF AQDTQAEGECSSRAERADMCPDAP PSQEVPEGPGAAP (SEQ ID NO:489). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in immune-related tissues such as thymus, macrophage, T cells and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders and infectious disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and infectious disease, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 271 as residues: Lys-9 to Arg-14, Met-38 to Asp-51.

- 5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, as well as the treatment and/or diagnosis of infectious disease.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 39

- 10 The translation product of this gene shares sequence homology with alpha-2 type I collagen which is thought to be important in tissue repair. (See, e.g., 211607.) Preferred polypeptide fragments comprise the amino acid sequence: PQLPSCGRPWP GTASVFQSH TQGPREDPDPCRAQGSAGTHCPISLSPPRQ (SEQ ID NO:490). Also preferred are the polynucleotide sequences encoding these polypeptide sequences.

- 15 This gene is expressed primarily in the brain and to a lesser extent in the kidney and thymus

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain, kidney, and immune disorders. Similarly, polypeptides and  
20 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, kidney, and immune disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,  
25 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 30 The tissue distribution and homology to alpha-2 type I collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tissue repair, and brain, kidney, immune disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 40

- 35 The translation product of this gene shares sequence homology with mini-collagen which is thought to be important in tissue repair tumor metastasis. (See Accession No. gnllPID1006976.) Preferred polypeptide fragments comprise the amino acid sequence: PGFRGPSGLGCSFFPRSLGRVLPPGCQRPGAHAD



SSPPPTP (SEQ ID NO:491). Also preferred are polynucleotides encoding this polypeptide fragment.

This gene is expressed in ovarian cancer and to a lesser extent in dendritic cells and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumor metastasis and tissue repair. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumor metastasis and tissue repair, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 273 as residues: Asn-2 to His-11.

The tissue distribution and homology to mini-collagen gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumor metastasis and tissue repair.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 41

This gene shares sequence homology with the HIV TAT protein. (See Accession No. 328416.) Preferred polypeptide fragments comprise the amino acid sequence: EDLKKPDPA~~S~~LAASCGEGK~~K~~RKACKNCTCGLAEELEKEK SREQMSSQPKSACGNCYL~~G~~DAFR~~C~~ASCPYLGMPAFKPG~~E~~KVLLS (SEQ ID NO:492); EDLKKPDPA~~S~~LAASCGEGK~~K~~RKACKNCTCGLAEELEKEK SREQMSSQPKSACGNCYL~~G~~DAFR~~C~~ASCPYLGMPAFKPG~~E~~KVLLSDSNLHD (SEQ ID NO:493); CGNCYL~~G~~DAFR~~C~~ASCPYLGMPAFKPG~~E~~KVLLSDS (SEQ ID NO:494); SCGEGK~~K~~RKACKNCTCGLAEELEKE (SEQ ID NO:495); SQPKSAC GNCYL~~G~~DAFR~~C~~ASC (SEQ ID NO:496); and REAGQNSERQYVS LSRD (SEQ ID NO:497). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in the infant brain and to a lesser extent in the breast and testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain, testes and breast disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, testes and breast disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 274 as residues: Pro-7 to Val-15.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of brain, testes and breast, and other related disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 42

This gene is expressed primarily in the infant brain, human cerebellum, and to a lesser extent in medulloblastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain related disorders and medulloblastoma and other brain cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain related disorders and brain cancers, including medulloblastoma, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 275 as residues: Thr-41 to Glu-47.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 44**

This gene is expressed primarily in the fetal brain, cerebellum and to a lesser extent in the placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal cell related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal cell related disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 277 as residues: Thr-20 to Gly-28.

The tissue distribution and homology to proline-rich protein genes indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal cell related disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 45**

The translation product of this gene shares sequence homology with precerebellin of human, which is thought to be important in synaptic physiology. (See Accession No. gi1180251.) It has been observed that cerebellin-like immunoreactivity is associated with Purkinje cell postsynaptic structures. Thus, it is likely that this gene also have synaptic activity. Preferred polypeptide fragments comprise the amino acid sequence: QEGSEPVLLERGECLVVCEPGRAAAGGPGGAALGEAPPGRVAFXAV RSHHHEPAGETGNGTSGAIYFDQVLVNEGGGFDRASGSFVAPVRGVYSFRFH VVKVYNRQTVQVSLMLNTWPVISAFANDPDVTREAAATSSVLLPLDPGDRVSLR LRRGXSTGW (SEQ ID NO:499). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in cerebellum and infant brain. By Northern analysis, a single transcript of 2.4 kb was observed in brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human brain related disorders, brain cancers, and medulloblastoma.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 43

The translation product of this gene shares sequence homology with a phosphotyrosine-independent ligand for the lck SH2 domain which is thought to be important in signal transduction related to phosphotyrosine-independent ligand for the lck SH2 domain. (See Accession No. gi1184951.) Preferred polypeptide fragments  
10 comprise the amino acid sequence: ESSGQARTLADPGPGWPRQQGMCFGSLT  
GLSTTPHGF LTVSAEADPRLIESLSQMLSMGFSDEGGWLTRLLQTKNYDIGAAL  
DTIQYSKH (SEQ ID NO:498). Also preferred are polynucleotide fragments encoding this polypeptide fragment. It is likely that this gene is a new member of a family of phosphotyrosine-independent ligands for the lck SH2 domains.

15 This gene is expressed primarily in the placenta and to a lesser extent in endothelial cells and neutrophil.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are  
20 not limited to, reproductive, cardiovascular, immune, and infectious diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular, reproductive, and immune system, and infectious diseases, expression  
25 of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the  
30 disorder.

The tissue distribution and homology to a phosphotyrosine-independent ligand for the lck SH2 domain indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cardiovascular, reproductive, and immune system diseases, as well as infectious diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 47**

The translation product of this gene shares sequence homology with a 12 kD nucleic acid binding protein of Feline calicivirus which is thought to be important in viral replication. (See Accession No. 59264)

5 This gene is expressed primarily in human cardiomyopathy and to a lesser extent in T helper cells, fetal brain and synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiomyopathy as well as viral infection. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 280 as residues: Trp-20 to Cys-26.

The tissue distribution in cardiomyopathy and homology to viral 12 kD nucleic acid binding protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of cardiomyopathy, including those caused by ischemic, hypertensive, congenital, valvular, or pericardial abnormalities.

25 The gene expression pattern may be the consequence or the cause for these conditions.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 48**

The translation product of this gene shares sequence homology with tumor necrosis factor related gene product which is thought to be important in tumor necrosis, bacterial and viral infection, immune diseases and immunoreactions.

This gene is expressed primarily in colon and to a lesser extent in ovarian and breast cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of colon, ovary or breast origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

not limited to, neuronal cell signal transduction and synaptic physiology. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal cell signal transduction and synaptic physiology expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to gene or gene family indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal cell related disorders.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 46

This gene is expressed in fetal liver and spleen, and to a lesser extent in bone marrow, umbilical vein, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the immune system, particularly hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis and immune disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 279 as residues: Asp-30 to Glu-57.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders.

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the colon, ovary and breast, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to Tumor necrosis factors indicates that polynucleotides and polypeptides corresponding to this gene are useful for intervention of cancers of colon, ovary and breast origins, because TNF family members are known to be involved in the tumor development.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 49

The translation product of this gene shares sequence homology with mucins, such as epithelial mucin, which is thought to be important in extracellular matrix functions such as protection, lubrication and cell adhesion (See for example Accession No. R68002). Preferred polypeptide fragments comprise the following amino acid sequence: PRSRPALRPGRQRPPSHSATSGVLRPRKKPDP (SEQ ID NO:500).

Also preferred are polynucleotide fragments encoding these polypeptide fragments. Moreover, this gene maps to chromosome 22q11.2-qter, and therefore, can be used as a marker in linkage analysis for chromosome 22.

This gene is expressed primarily in corpus colosum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors, especially of corpus colosum, as well as metastatic lesions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the corpus colosum and other solid tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to mucins indicates that polynucleotides and polypeptides corresponding to this gene are useful for serum tumor markers or immunotherapy targets because tumor cells have greatly elevated level of mucin expression and shed the molecules into the epithelial tissues.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 50

This gene is expressed primarily in CD34 depleted buffy coat cord blood and primary dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic disorders and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For  
15 a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard  
20 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in CD34 depleted buffy coat cord blood and primary dendritic cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders.  
25 Secreted or cell surface proteins in the above tissue distribution often are involved in cell activation (e.g. cytokines) or molecules involved in cell surface activation.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 51

The translation product of this gene shares sequence homology with Interferon  
30 induced 1-8 gene encoded polypeptide which is thought to be important in binding to retroviral rev responsive element. Preferred polypeptide fragment comprise the following amino acid sequences: MTLITPSXKLTFXKGNKSWSSRACSSTLVDP (SEQ ID NO:501). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

35 This gene is expressed primarily in CD34 positive cells and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a



biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, retroviral infection, such as AIDS, and other immune disorders.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 284 as residues: Gln-51 to Trp-62.

The tissue distribution and homology to interferon induced gene 1-8 indicates that polynucleotides and polypeptides corresponding to this gene are useful for intervention of retroviral infection including HIV. The factor may be involved in viral stability or viral entry into the cells. Alternatively, the virus/factor complex may elicit the cellular immune reaction.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 52

This gene shares sequence homology to immunoglobulin lambda chain (See Accession No. 2865484). Therefore it is likely that this gene has activity similar to an immunoglobulin lambda chain. Preferred polypeptide fragments comprise the following amino acid sequence: GHPSPALSIAPSDGSQPCDEVYPYGEAHVTRYCKKPLTNS HLETEAQSSSL (SEQ ID NO:502). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in Hodgkin's lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, Hodgkin's lymphoma and other immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 285 as residues: Pro-27 to Thr-32.

The tissue distribution in Hodgkin's lymphoma and the sequence homology indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of Hodgkin's lymphoma, since the elevated expression and secretion by the tumor mass may be indicative of tumors of this type. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 53

This gene has extensive homology to cDNA for Homo sapiens mRNA for the ISLR gene (See Accession No. AB003184). This protein is considered to be a new member of the Ig superfamily and contains a leucine-rich repeat (LRR) with conserved flanking sequences and a C2-type immunoglobulin (Ig)-like domain. These domains are important for protein-protein interaction or cell adhesion, and therefore it is possible that the novel protein ISLR may also interact with other proteins or cells. The ISLR gene was mapped on human chromosome 15q23-q24 by fluorescence in situ hybridization (See Medline Article No. 97468140). Homology to the ISLR gene has been confirmed by another independent group as well (See Accession No. Hs.102171)

This gene is expressed in a number of tissues including human retina, heart, skeletal muscle, prostate, ovary, small intestine, thyroid, adrenal cortex, testis, stomach, spinal cord, fetal lung and fetal kidney tissues, colon, tonsil and stomach cancer, and to a lesser extent in endometrial stromal cells treated with estradiol, breast tissue, synovium, lymphoma, and number of other tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of colon, ovary and breast origins. However, due to the wide range of expression in various tissues, protein may play a vital role in the development of cancer in other tissues as well, not just those mentioned above. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the colon, ovary and breast, expression of this gene at significantly higher or lower levels may be routinely

detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Additionally, this gene maps to chromosome 15q23-q24, and therefore, can be used as a marker in linkage analysis for chromosome 15.

The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 54

Gene has homology to multidrug resistance gene 1 (See Accession No. P06795). Preferred polynucleotide fragments comprise the following sequence:  
 GCTTCGTGTCCAACCCTCTTGCCCTTCGCCTGTGTGCCTGGAGCCAGTCCCA  
 CCACGCTCGCGTTTCCTCCTGTAGTGCTCACAGGTCCCAGCACCGATGGCA  
 TTCCCTTTGCCCTGAGTCTGCAGCGGGTCCCTTTTGTGCTTCCTTCCCCTCA  
 GGTAGCCTCTCTCCCCCTGGGCCACTCCCGGGGGTGAGGGGGTTACCCCTT  
 CCCAGTGTTTTTTATTCCTGTGGGGCTACCCCAAAGTATTAAAAGTAGCTTT  
 GTAA (SEQ ID NO:503). Also preferred are polypeptide fragments encoded by these polynucleotide fragments.

This gene is expressed primarily in lung, esophagus, leukemia (Jurkat cells) and breast cancers and to a lesser extent in macrophages treated with GM-CSF fetal tissues and wide range of tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer of wide range of origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the solid tumors, lung and leukemia, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Furthermore, due to the high expression level in lung tissue and the proposed function of the multidrug resistance protein 1 gene as the efflux pump responsible for low-drug accumulation in multidrug-resistant cells, protein as well mutants thereof, may also be beneficial as a target for gene therapy, particularly for the chronic patient. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 287 as residues: Met-1 to Lys-16.

The tissue distribution in wide range of cancers and fetal tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of cells in active proliferation, such as cancers. The gene products may be used for cancer markers or immunotherapy target.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 55

This gene maps to the X chromosome.

This gene is expressed primarily in the brain and to a lesser extent in the developing embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disease states and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders, including sex-linked disorders, of the above tissues or cells, particularly of the neurological, developmental systems, and cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Moreover, this gene maps to the X chromosome, and therefore, may be used as a marker in linkage analysis for this chromosome.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Klinefelter's, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental

disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 56

- 5 The translation product of this gene shares sequence homology with paxillin which is thought to be important in mediating signal transduction from growth factor receptors to the cytoskeleton. Preferred polynucleotide fragments comprise the following sequence: TGGCTCACTGTCTTACAATCACTGCTGTGGAATCATGA  
TACCACTTTTAGCTCTTTGCATCTTCCTTCAGTGTATTTTGTGTTTTCAAGAGG  
10 AAGTAGATTTTAACTGGACAACCTTTGAGTACTGACATCATTGATAAAATAAACT  
GGCTTGTGGTTTCAA (SEQ ID NO:506). Also preferred are polypeptide fragments encoded by these polynucleotide fragments. More preferably, polypeptide fragments comprise the amino acid sequence: LDELMAHLTEMQAKVAVRAD  
AGKKHLPDKQDHKASLDSMLGGLEQELQDLGIATVPKGHCASCQKPIAGKVI  
15 HALGQSWHPEHFVCTHCKEEIGSSPFFERSGLXYCPNDYHQLFSPRCAYCAAP  
ILDKVLTAMNQTWHPEHFFCSHCGEVFGAEGFHEKDKKPYCRKDFLAMFSPK  
CGGCNRPVLENYLSAMDTVWHPECFVCGDCFTSFSTGSFFELDGRPFCELYH  
HRRGTLCHGCGQPITGRCSAMGYKFHPEHFVCAFCLTQLSKGIFREQNDKTY  
CQPCFNKLF (SEQ ID NO:507); KASLDSMLGGLEQELQDLGIATVPKGHC  
20 ASCQKPIAGKVIHAL (SEQ ID NO:508); CPNDYHQLFSPRCAYCAAPILDKVL  
TAMNQTWHPEHFFCSHCGEVFGAEG (SEQ ID NO:509); DKKPYCRKDFLAM  
FSPKCGGCNRPVLENYLSAMDTVWHPECFVCGDCFTSFSTGSFFELDGRPFCE  
L (SEQ ID NO:510); CGQPITGRCSAMGYKFHPEHFVCAFCLTQLSKGIFRE  
QNDKTYCQ (SEQ ID NO:511). Polynucleotide fragments encoding these preferred  
25 polypeptide fragments are also contemplated.

This gene is expressed primarily in brain, and to a lesser extent in the developing embryo.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disease states and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and  
35 nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or

cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Moreover, since this gene shares homology with a gene that maps to chromosome 11, (See Accession No.T87404), gene as well as its translated product may be used for linkage analysis on chromosome 11.

The tissue distribution and homology to paxillin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and or detection of disease states associated with abnormal signal transduction in brain and/or the developing embryo. This would include treatment or detection of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder and also in the treatment and or detection of embryonic development defects.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 57

This gene is expressed primarily in fetal spleen, brain, and to a lesser extent in six week old embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders, neurological disorders, and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and developmental systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 290 as residues: Arg-28 to Gly-34.

The expression of this gene in fetal spleen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/detection of immune disorders such as arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. In addition the expression of this gene in the early embryo, indicates a key role in embryo development and hence the gene or gene product could be used in the treatment and or detection of embryonic development defects. This would include

treatment or detection of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntinton's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder and also in the treatment and or detection of embryonic development defects.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 58

The translation product of this gene shares sequence homology with the gene disrupted in the neurodegenerative disease dentatorubal-pallidoluysian atrophy. Moreover a long open reading frame exists in an alternative frame. Preferred polypeptide fragments

10

comprise the following:

15

MGSSQSVEIPGGGTEGYHVLRVQENSPGHRAGLEPFFDFIVSINGSRLNKDND  
 TLKDLLKXNVEKPKV KMLIYSSKTLELRETSVTPSNLWGGQGLLGVSIRFCSFD  
 GANENVVHVLEVESNSPAALAGLRPHSDYIIGADTMNESEDLFSLIETHEAKP  
 LKLYVYNTD TDNCREVIITPNSAWGGEGLGCGIGYGYLHRIPTRPFEEGKKIS  
 LPGQMAGTPITPLKDGFTVQLSSVNPPSLSPPGTTGIEQSLTGLSISSTPPAVSS  
 VLSTGVPTVPLLPPQVNQSLTSVPPMNPATTL PGLMPLPAGLPNLPNLNLNLP  
 PHIMPGVGLPELVNPGLPPLPSMPPRNLP GIAPLPLPSEFLPSFPLVPESSSAASS  
 GELLSSLPPTSNAPSDPATTTAKADAASSLTVDVTPPTAKAPTTVEDRVGDSTPV  
 SEKPVSAAVDANASESP (SEQ ID NO:512); SVEIPGGGTEGYHVLRVQENSPGH  
 20 RAGLEPFFDFIVSINGSRLNKDNDTLKDLLKXNVEKPKV KMLIYSSKTLELRETS  
 VTPSNLWGGQGLLGVSIRFCSFDGANENVVH (SEQ ID NO:513); ESNSPA  
 LAGLRPHSDYIIGADTMNESEDLFSLIETHEAKPLKLYVYNTD TDNCREVIITP  
 NSAWGGEGLGCGIGYGYLHRIPTRPFEEGKKISLPGQMAGTPITPLKDGFTV  
 QLSSVNPPSLSPPGTTGIEQSLTG LSIS (SEQ ID NO:514); RIPTRPFEEGKKI  
 25 SLPGQMAGTPITPLKDGFTVQLSSVNPPSLSPPGTTGIEQSLTGLSISSTPPAVS  
 SVLSTGVPTVPLLPPQVNQSLTSVPPMNPATTL PGLMPLPAGLPNLPNLNLNLP  
 APHIMPGVGLPELVNPGLPPLPSMPPRN (SEQ ID NO:516); PGLPPLPSMPPRN  
 LPGIAPLPLPSEFLPSFPLVPESSSAASSGELLSSLPPTSNAPSDPATTTAKADAA  
 SSLTVDVTPPTAKAPTTVEDRVGDSTPVSEKPVSAAVDAN (SEQ ID NO:517).

30

This gene is expressed primarily in prostate cancer, and to a lesser extent in the pineal glands and in fetal lung.

35

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological conditions and pulmonary disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For

a number of disorders of the above tissues or cells, particularly of the nervous, pulmonary, and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or  
5 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 291 as residues: Asn-9 to Leu-14.

The abundance of this gene in the pineal gland and its homology to a gene  
10 disrupted in the neurodegenerative disease state Dentatorubral-pallidoluysian atrophy indicates that this gene may be useful in the treatment and/or detection of other neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. The abundance of this gene in fetal  
15 lung would suggest that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung; that it may also be involved in predisposition to certain pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis; and thus the gen or the gene protein encoded by the gene could be used in the detection and/or treatment of these  
20 pulmonary disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 59

This gene is expressed primarily in the developing embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as  
25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of  
30 the above tissues or cells, particularly of the developmental system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the  
35 expression level in healthy tissue or bodily fluid from an individual not having the disorder.



The expression of this gene primarily in the embryo, indicates the gene plays a key role in embryo development and that the gene or the protein encoded by the gene could be used in the treatment and or detection of developmental defects in the embryo or in infants.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 60

This gene displays homology to nestin, an intermediate filament protein, the expression of which correlates with the proliferation of Central Nervous System progenitor cells and that is useful in the identification of brain tumors. This gene maps  
10 to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. AA527348).

This gene is expressed primarily in kidney and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
15 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, renal disorders and neurodegenerative conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the excretory and  
20 nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
25 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 293 as residues: Thr-128 to Asn-135.

The tissue distribution and homology to nestin indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and/or treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease,  
30 Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, its abundance in kidney indicates that it is useful in the treatment and detection of acute renal failure and other disease states associated with the kidney.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 61

Gene shares homology with the latrophilin-related protein 1 precursor as well as the calcium-independent alpha-latrotoxin receptor. Preferred polypeptide fragments

comprise the following amino acid sequence:

IYKVFRHTAGLKPEVSCFENIRSCARXXXXXXXXXXXXXWIFGVLHVHVSVV  
TAYLFTVSNAFQGMFIFLFLCVLSRKIQEEYYRLFKNVPCC (SEQ ID NO:518);  
WIFGVLHVHVSVV TAYLFTVSNAFQGMFIFLFLCVLSRKIQEEYYRLFKNVPC

- 5 C (SEQ ID NO:519): Also preferred are polynucleotide fragments encoding these polypeptide fragments. (See Accession No. 2213659) The translation product of this gene shares sequence homology with CD 97, a seven transmembrane bound receptor.

This gene is expressed primarily in infant brain and in endothelial cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For  
15 a number of disorders of the above tissues or cells, particularly of the neurological and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the  
20 standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 294 as residues: Lys-13 to Leu-21.

- The tissue distribution of this gene suggest that it may be useful in the detection and/or treatment of neurodegenerative disease states and behavioral disorders such as  
25 Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder, while its expression in hematopoietic cell types indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, asthma and immunodeficiency diseases.

30

## FEATURES OF PROTEIN ENCODED BY GENE NO: 62

This gene is expressed primarily in fetal liver and fetal spleen.

- Therefore, polynucleotides and polypeptides of the invention are useful as  
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematological and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 295 as residues: Ser-91 to Lys-98.

The tissue distribution of this gene fetal liver and spleen indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, leukemia, asthma and immunodeficiency diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 63

Gene shares homology with human serum amyloid protein. Preferred polypeptide fragments comprise the following amino acid sequence:  
ALTRIPPGDWVINVTAVSFAGKTTARFFHSSPPSLGDAQARTDPGHQRRD (SEQ ID NO:520) (See Accession No. W13671). Also preferred are polynucleotide fragments encoding these polypeptide fragments. This gene maps to chromosome 9, and therefore, may be used as a marker in linkage analysis for chromosome 9 (See Accession No. AA004342).

This gene is expressed primarily in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in fetal liver-spleen indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, leukemia, asthma, and immunodeficiency diseases.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 64

This gene maps to chromosome 3, and therefore, may be used as a marker in linkage analysis for chromosome 3 (See Accession No. AA219669).

This gene is expressed specifically in the brain.

Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disease states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of  
15 the above tissues or cells, particularly of the neurological systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the  
20 expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's  
25 Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 65

Gene shares homology with a yeast protein. Preferred polypeptide fragments  
30 comprise the following amino acid sequence: LQEVNITLPENSVWYERYKFDIP VFHL (SEQ ID NO:521). Also preferred are polynucleotide fragments encoding these polypeptide fragments. (See Accession No. 1332638)

This gene is expressed primarily in fetal tissue (fetus and fetal liver).

Therefore, polynucleotides and polypeptides of the invention are useful as  
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, liver disorders and cancers (e.g. hepatoblastoma). Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 298 as residues: Asn-59 to Glu-64.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 66

Gene has homology with a B-cell surface antigen which may indicate gene plays a role in the immune response, including, but not limited to disorders and infections of the immune system. Preferred polynucleotide fragments comprise the following sequence: TAGCATGTAGCCAGTCTGAATAACNTATAAGGACAAAGTGGAGTC CACGCGTGCGGCCGTCTAGACTAGTGGATCCCCGGCTGCAGGATTCGGC ACGAG (SEQ ID NO:523). Also preferred are polypeptide fragments encoded by these polynucleotide fragments (See Accession No.T94535). Additionally, this gene shares homology with an interferon-gamma receptor. Preferred polypeptide fragments also comprise the following amino acid sequence: MQGSGSQFRACLLCLCFSCPC SPGGPRWNSRQGRRFPKTCRAISQNLVFKYKTFPCVRYMQPHRSSLCLHFTS YVFILSTWGLRITYSTDLLKKKKNSRGGPVPIRPKS (SEQ ID NO:522); MQGSGSQFRACLLCLCFSCPCSPGGPRWNSRQGRRFPKTCRAISQNLVFK (SEQ ID NO:524); PVRYMQPHRSSLCLHFTSYVFILSTWGLRITYSTDLLKKKK NSRGGPVPIRPKS (SEQ ID NO:525); and GEEQRDCSLGWRGVGMRATHCQAA RMFVLFSLPKYAGL (SEQ ID NO:526). Also preferred are polynucleotide fragments encoding these polypeptide fragments

This gene is expressed primarily in T-cells and gall bladder.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological disorders and conditions (immunodeficiencies, cancer, leukemia, hematopoiesis). Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and digestive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 299 as residues: Thr-41 to Gly-52.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune disorders, immunosuppressive (transplantation) and immunodeficiencies (e.g. AIDS), inflammation and hematopoietic disorders. The expression of this gene in gall bladder would suggest a possible role for this gene product in digestive disorders, particularly of the pancreas.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 67

This gene maps to chromosome 11, and therefore, may be used as a marker in linkage analysis for chromosome 11 (See Accession No. AA011622).

This gene is expressed primarily in a variety of fetal and developmental tissues (e.g. fetal spleen, infant brain).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental, immune or neurological abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing immune and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 300 as residues: Ser-38 to Ser-43.

5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for developmental abnormalities or fetal deficiencies. The detection in infant brain would suggest a role in neurological disorders (both developmental and neurodegenerative conditions of the brain and nervous system, behavioral disorders, depression, schizophrenia, Alzheimer's disease, Parkinson's  
10 disease, Huntington's disease, mania, dementia). In addition, the detection in spleen would similarly suggest a role in detection and treatment of immunologically mediated disorders (e.g. immunodeficiency, inflammation, cancer, wound healing, tissue repair, hematopoiesis).

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 68

This gene is expressed primarily in spleen, T-cells, and fetal heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are  
20 not limited to, immunological deficiencies, including AIDS and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and cardiovascular systems, expression of this gene at significantly higher  
25 or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, autoimmune disorders, immunodeficiencies (e.g. AIDS), immuno-suppressive conditions (transplantation) and hematopoietic disorders. The expression in fetal heart indicates that polynucleotides and  
35 polypeptides corresponding to this gene are useful for the treatment and diagnosis of cardiovascular disorders (e.g. heart disease, restenosis, atherosclerosis, stroke, angina, thrombosis).

**FEATURES OF PROTEIN ENCODED BY GENE NO: 69**

Gene shares homology with a human collagen protein. Preferred polypeptide fragments comprise the following amino acid sequence:

- 5 MPRKTSKCRQLLCSGASRNADTAARQSTCSSHRPPGKIPSLGPRRXPGCXSV  
SSRGEQSTGSPAAPRCGRDAHRGLPGGAAMTPGDTWASFNPRAGHSKSQGE  
QGESSGASRQDRHPVSHWVERQREAWGAPRSSSAGGVKVAATTEREPEFKIK  
TGKA (SEQ ID NO:527); CSGASRNADTAARQSTCSSHRPPGKIPSLGPRRXPG  
CXSVPSRGEQSTGSPAAPRCGRDAHRGLPGGAAMTPGDTWASFNPRAGHS  
10 (SEQ ID NO:528); QGEGQESSGASRQDRHPVSHWVERQREAWGAPRSSSAGG  
VKVAATTEREPEFKIKTGKA (SEQ ID NO:529) (See Accession No. 124886). Also  
preferred are polynucleotide fragments encoding these polypeptide fragments

This gene is expressed primarily in fetal heart.

- Therefore, polynucleotides and polypeptides of the invention are useful as  
15 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
not limited to, cardiovascular disorders. Similarly, polypeptides and antibodies directed  
to these polypeptides are useful in providing immunological probes for differential  
identification of the tissue(s) or cell type(s). For a number of disorders of the above  
20 tissues or cells, particularly of the cardiovascular system, expression of this gene at  
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,  
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial  
fluid or spinal fluid) or another tissue or cell sample taken from an individual having  
such a disorder, relative to the standard gene expression level, i.e., the expression level  
25 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred  
epitopes include those comprising a sequence shown in SEQ ID NO: 302 as residues:  
Pro-32 to Ser-39.

- The tissue distribution indicates that polynucleotides and polypeptides  
corresponding to this gene are useful for the treatment and diagnosis of cardiovascular  
30 disorders (e.g. heart disease, restenosis, atherosclerosis, stroke, angina, thrombosis).

**FEATURES OF PROTEIN ENCODED BY GENE NO: 70**

- The translation product of this gene shares sequence homology with a chicken  
single-strand DNA-binding protein. Preferred polypeptide fragments comprise the  
35 following amino acid sequence:  
MSPRYPGGPRPPLRIPNQALGGVPGSQPLLPSGMDPTRQQGHPNMGGPMQRM  
TPPRGMVPLGPQNYGGAMRPLNALGGPGMPGMNMGPGGGRPWPNTNAN



SIPYSSASPGNYVGPPGGGGPPGTPIMPSPADSTNSGDNMYTLMNAVPPGPNR  
 PNFPMPGSDGPMGGLGGMESHMHMNGSLGSGDMDISISKNPNNMSLSNQPG  
 GTPRDDGEMGGNFLNPFQSESYSPSMTMSV (SEQ ID NO:530); MSPRYPGG  
 PRPPLRIPNQALGGVPGSQPLLPSGMDPTRQQGHPNMGGPMQRMTPPRGMVP  
 5 LGPQNYGGAMRPPLNALGGPGMPGMNMGPGGGRPWPNPNTNANSIPYSSASP  
 GNY (SEQ ID NO:531); LNALGGPGMPGMNMGPGGGRPWPNPNTNANSIPYSS  
 ASPGNYVGPPGGGGPPGTPIMPSPADSTNSGDNMYTLMNAVPPGPN (SEQ ID  
 NO:532); GPMGGLGGMESHMHMNGSLGSGDMDISISKNPNNMSLSNQPGTPR  
 DDGEMGGNFLNPFQSESYSPSMTMSV (SEQ ID NO:533); TCEHSSEAKAFHDY  
 10 (SEQ ID NO:534). Also preferred are polynucleotide fragments encoding these  
 polypeptide fragments. (See Accession No. 1562534)

This gene is expressed primarily in placenta and to a lesser extent in the fetal heart and a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 15 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions, which include, but are  
 not limited to, developmental abnormalities, fetal deficiencies, and particularly of the  
 cardiovascular system. Similarly, polypeptides and antibodies directed to these  
 polypeptides are useful in providing immunological probes for differential identification  
 20 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,  
 particularly of the reproductive system, expression of this gene at significantly higher or  
 lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded  
 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or  
 another tissue or cell sample taken from an individual having such a disorder, relative to  
 25 the standard gene expression level, i.e., the expression level in healthy tissue or bodily  
 fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
 corresponding to this gene are useful for the detection and treatment of developmental  
 abnormalities or fetal deficiencies, ovarian and other endometrial cancers, reproductive  
 30 dysfunction, cardiovascular disorders, and pre-natal disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 71

This gene is expressed primarily in fetal liver and to a lesser extent in the breast  
 and testes.

35 Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, liver disorders (including hepatoblastomas) and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). The expression in testes and breast indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of endocrine and reproductive disorders (e.g. sperm maturation, milk production, testicular and breast cancers).

## 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 72

This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. W93595).

This gene is expressed primarily in smooth muscle and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiovascular and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of restenosis, atherosclerosis, stroke, angina, thrombosis, wound healing and other conditions of heart disease. In addition, the expression in brain would suggest that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of developmental, degenerative and behavioral conditions of the brain and nervous system (e.g. schizophrenia, depression, Alzheimer's disease, Parkinson's disease, Huntington's disease, mania, dementia, paranoia, addictive behavior and sleep disorders).

### FEATURES OF PROTEIN ENCODED BY GENE NO: 73

Gene shares homology with human stromalin-2. Preferred polypeptide fragments comprise the following amino acid sequence:

QAFVLLSDLLLIFSPQMIVGGRDFLRPLVFFPEATLQSELASFLMDHVFQPGDL  
GSGA (SEQ ID NO:535); ACSYLLCNPEFTFFSRADFARSQVLDLLTDRFQQE  
LEELLQVG (SEQ ID NO:536); QKQLSSLRDRMVAFCCLCQSCLSVDTEIQEQV  
ST (SEQ ID NO:537); QVILPALTLYYFSILWTLTHISKSDAS (SEQ ID NO:538);  
STHDLTRWELYECCQLLQKAVDTGXVPHQV (SEQ ID NO:539). Also preferred

are polynucleotide fragments encoding these polypeptide fragments (See Accession No. R65208 ) This gene maps to chromosome 7, and therefore, may be used as a marker in linkage analysis for chromosome 7 (See Accession No. D52585).

This gene is expressed primarily in the brain (infant brain, adult brain, pituitary, cerebellum, hippocampus, schizophrenic hypothalamus, amygdala).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental and neurodegenerative diseases of the brain and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those

comprising a sequence shown in SEQ ID NO: 306 as residues: Thr-25 to Lys-36, Lys-55 to Ser-63.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of developmental, degenerative and behavioral conditions of the brain and nervous system (e.g. schizophrenia, depression, Alzheimer's disease, Parkinson's disease, Huntington's disease, mania, dementia, paranoia, addictive behavior and sleep disorders).

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 74

This gene is expressed primarily in the hypothalamus of a human suffering from schizophrenia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the CNS particularly schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS, such as schizophrenia expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 307 as residues: Gly-38 to Ala-44.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of schizophrenia and other disorders involving the CNS.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 75

Preferred polypeptides of the invention comprise the following amino acid sequence encoded by this gene:

LAVSTSFICCADISTALPLGSSRPAPAPRHREHEHGHQARPPRLLXTSLMPLSTP  
AAAQLLWTQLTPMGGRPGGRHSPPTLHTGPRALPPGPPHPSLHVAALSLLR

(SEQ ID NO:540). Polynucleotides encoding such polypeptides are also provided.

This gene is expressed primarily in endometrial tumor and to a lesser extent in amniotic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive and immune disorders particularly cancers of those systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 308 as residues: Ser-3 to Arg-9.

The tissue distribution indicates that the protein products of this gene are useful for study and treatment of immune and reproductive disorders particularly cancers of those systems.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 76

This gene is expressed primarily in kidney cortex and to a lesser extent in early stage human brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, renal disorders such as renal cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 309 as residues: Gly-38 to Gly-45, Gly-47 to Gly-52, Pro-92 to Lys-110.

The tissue distribution indicates that the protein products of this gene are useful for study, treatment and diagnosis of renal diseases such as cancer of the kidney.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 77**

This gene is expressed primarily in kidney medulla.

Therefore, polynucleotides and polypeptides of the invention are useful as  
5 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic and renal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of  
10 the above tissues or cells, particularly of the metabolic and renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,  
15 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for study, treatment and diagnosis of metabolic and renal diseases and disorders.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 78**

This gene is expressed in chronic synovitis and microvascular endothelium.

Therefore, polynucleotides and polypeptides of the invention are useful as  
25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, arthritis and atherosclerosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular and skeletal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain  
30 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of arthritic and other inflammatory diseases as well as cardiovascular diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 79**

This gene is expressed in resting T-cells and activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as  
5 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
not limited to, immune disorders. Similarly, polypeptides and antibodies directed to  
these polypeptides are useful in providing immunological probes for differential  
10 identification of the tissue(s) or cell type(s). For a number of disorders of the above  
tissues or cells, particularly of the immune system, expression of this gene at  
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,  
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial  
fluid or spinal fluid) or another tissue or cell sample taken from an individual having  
15 such a disorder, relative to the standard gene expression level, i.e., the expression level  
in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful  
for the study and treatment of immune diseases such as inflammatory conditions.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 80**

20 This gene is expressed in a variety of immune system tissues, e.g., neutrophils,  
T-cells, and TNF induced epithelial and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
25 not limited to, infectious and immune disorders. Similarly, polypeptides and antibodies  
directed to these polypeptides are useful in providing immunological probes for  
differential identification of the tissue(s) or cell type(s). For a number of disorders of  
the above tissues or cells, particularly of the immune and vascular systems, expression  
of this gene at significantly higher or lower levels may be routinely detected in certain  
30 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,  
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an  
individual having such a disorder, relative to the standard gene expression level, i.e.,  
the expression level in healthy tissue or bodily fluid from an individual not having the  
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID  
35 NO: 313 as residues: Met-1 to Trp-6.

The tissue distribution indicates that the protein products of this gene are useful  
for study and treatment of infectious diseases, immune and vascular disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 81**

This gene is expressed in activated neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and other immune conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for study and treatment of immune disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 82**

This gene is expressed in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory and other immune conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 315 as residues: Ala-83 to Thr-91.

The tissue distribution indicates that the protein products of this gene are useful for study and treatment of immune disorders.



**FEATURES OF PROTEIN ENCODED BY GENE NO: 83**

This gene is expressed in human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as  
5 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
not limited to, inflammation and immune disorders. Similarly, polypeptides and  
antibodies directed to these polypeptides are useful in providing immunological probes  
for differential identification of the tissue(s) or cell type(s). For a number of disorders  
10 of the above tissues or cells, particularly of the immune and inflammatory system,  
expression of this gene at significantly higher or lower levels may be routinely detected  
in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,  
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from  
an individual having such a disorder, relative to the standard gene expression level, i.e.,  
15 the expression level in healthy tissue or bodily fluid from an individual not having the  
disorder.

The tissue distribution indicates that the protein products of this gene are useful  
for diagnosis and treatment of disorders of the inflammatory and immune systems.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 84**

This gene is expressed in human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
25 not limited to, disorders of the inflammatory and immune systems. Similarly,  
polypeptides and antibodies directed to these polypeptides are useful in providing  
immunological probes for differential identification of the tissue(s) or cell type(s). For  
a number of disorders of the above tissues or cells, particularly of the inflammatory and  
immune systems, expression of this gene at significantly higher or lower levels may be  
30 routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily  
fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or  
cell sample taken from an individual having such a disorder, relative to the standard  
gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
individual not having the disorder.

35 The tissue distribution indicates that the protein products of this gene are useful  
for diagnosis and treatment of disorders of the immune and inflammatory systems.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 85**

This gene is expressed in activated neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as  
5 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
not limited to, inflammation and immune system diseases. Similarly, polypeptides and  
antibodies directed to these polypeptides are useful in providing immunological probes  
for differential identification of the tissue(s) or cell type(s). For a number of disorders  
10 of the above tissues or cells, particularly of the immune system and inflammatory  
system, expression of this gene at significantly higher or lower levels may be routinely  
detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,  
serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample  
taken from an individual having such a disorder, relative to the standard gene  
15 expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful  
for diagnosis and treatment of diseases of the inflammatory and immune systems.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 86**

This gene is expressed in activated neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
25 not limited to, inflammation and immune system disorders. Similarly, polypeptides and  
antibodies directed to these polypeptides are useful in providing immunological probes  
for differential identification of the tissue(s) or cell type(s). For a number of disorders  
of the above tissues or cells, particularly of the inflammatory and immune system,  
expression of this gene at significantly higher or lower levels may be routinely detected  
30 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,  
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from  
an individual having such a disorder, relative to the standard gene expression level, i.e.,  
the expression level in healthy tissue or bodily fluid from an individual not having the  
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID  
35 NO: 319 as residues: Met-1 to Gly-6, Gly-32 to Pro-43, Leu-55 to Gln-60.

The tissue distribution indicates that the protein products of this gene are useful  
for diagnosis and treatment of disorders of the immune and inflammatory system.

# FEATURES OF PROTEIN ENCODED BY GENE NO: 87

In specific embodiments, polypeptides of the invention comprise the sequence:

EQVLALLWPRFELILEMNVQSVRSTDPQRLGGLDTRPHYITRRYAEFSSALVSIN  
 5 QTIPNERTMQLLGQLQVEVENFVLRVAAEFSSRKEQLVFLINNYDMMLGVLME  
 RAADDSSKEVESFQQLLNARTQEFIEELLSPFFGGLVAFVKEAEALIERGQAERLR  
 GEEARVTQLIRGFGSSWKSSVESLSQDVMRSFTNFRNGTSIIQG (SEQ ID  
 NO:541), ALLKYRFFYQFLLGNERATAKEIRDEYVETLSKIYLSYYRSYLGRMLMK  
 VQYEEVAEKDDLGMGVEDTAKKGFXXSKPSRSRNTIFTLGTRGSVISPTELEAPILV  
 10 PHTAQR (SEQ ID NO: 542); EQRYPFEALFRSQHYXLLDNSCREYLFICEFFVVS  
 GPXAHDLFHAVMGRTLSMTLKHLDSYLADCYDAIAVFLCIHIVLRFRNIAAKRD  
 VPALDRYW (SEQ ID NO:543), GGLDTRPHYITRRYAEFSSALVSINQ (SEQ ID  
 NO:544); SRKEQLVFLINNYDMMLGVL (SEQ ID NO: 545) and/or ALLKYRFFY  
 QFLLGNERATAKEIRDEYVETLSKIYLSYYRSYLGRMLMKVQYEEVAEKDDLGMG  
 15 VEDTAKKGFXXSKPSLSRNTIFTLGTRGSVISPTELEAPILVPHTAQRXEQRYPF  
 EALFRSQHYXLLDNSCREYLFICEFFVVS GPXAHDLFHAVMGRTLSMTLKHL  
 SYLADCYDAIAVFLCIHIVLRFRNIAAKRDVPALDRYWEQVLALLWPRFELILEM  
 NVQSVRSTDPQRLGGLDTRPHYITRRYAEFSSALVSINQTIPNERTMQLLGQLQV  
 EVENFVLRVAAEFSSRKEQLVFLINNYDMMLGVL MERAADDSSKEVESFQQLLN  
 20 ARTQEFIEELLSPFFGGLVAFVKEAEALIERGQAERLRGEEARVTQLIRGFGSSW  
 KSSVESLSQDVMRSFTNFRNGTS (SEQ ID NO:546). Polynucleotides encoding  
 these polypeptides are also encompassed by the invention. The translation product of  
 this gene shares sequence homology with suppressor of actin mutation which is thought  
 to be important in mutation suppression.

25 This gene is expressed primarily in fetal liver and to a lesser extent in a variety  
 of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions, which include, but are  
 30 not limited to, liver and mutations. Similarly, polypeptides and antibodies directed to  
 these polypeptides are useful in providing immunological probes for differential  
 identification of the tissue(s) or cell type(s). For a number of disorders of the above  
 tissues or cells, particularly of the liver or cancer, expression of this gene at  
 significantly higher or lower levels may be routinely detected in certain tissues (e.g.,  
 35 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial  
 fluid or spinal fluid) or another tissue or cell sample taken from an individual having  
 such a disorder, relative to the standard gene expression level, i.e., the expression level

in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 320 as residues: Val-53 to Arg-60, Thr-88 to Thr-94, Ala-142 to Ser-150, Gly-188 to Glu-196, Gly-208 to Ser-214, Thr-227 to Gly-232, Lys-279 to Phe-285.

5 The tissue distribution and homology to suppressor of actin mutation suggest that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and of liver disorder or cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 88

10 This gene maps to chromosome 9, and therefore can be used in linkage analysis as a marker for chromosome 9. In specific embodiments, polypeptides of the invention comprise the sequence:

YEGKEFDYVFSIDVNEGGPSYKLPYNTSDDPWLTAYNFLQKNDLNPMFLDQVA  
KFIIDNTKGQMLGLGNPSFSDPFTGGGRYVPGSSGSSNTLPTADPFTGAGRYV  
15 PGSASMGTMMAGVDPFTGNSAYRSAASKTMNIYFPKKEAVTFDQANPTQILGK  
LKELNGTAPEEKKLTEDDLILLEKILSLICNSSSEKPTVQQLQILWKAINCPEDIV  
FPALDILRLSIKHPSVNENFCNEKEGAQFSSHLINLLNPKGKPANQLLALRTFC  
NCFVGQAGQKLMMSQRESLM SHAIELKSGSNKNI (SEQ ID NO: 547);

HIALATLALNYSVCFHKD (SEQ ID NO: 548); HNIEGKAQCLSLISTILEVVQ

20 DLEATFRLVALGTLISDDSNVQLAKS (SEQ ID NO: 549); LGVDSQIKKYSS  
VSEPAKVSECCRFILNLL (SEQ ID NO: 550); and/or YEGKEFDYVFSIDVNEGGPS  
YKLPYNTSDDPWLTAYNFLQKNDLNPMFLDQVAKFIIDNTKGQMLGLGNPSFS  
DPFTGGGRYVPGSSGSSNTLPTADPFTGAGRYVPGSASMGTMMAGVDPFTGN  
SAYRSAASKTMNIYFPKKEAVTFDQANPTQILGKLKELNGTAPEEKKLTEDDLI  
25 LLEKILSLICNSSSEKPTVQQLQILWKAINCPEDIVFPALDILRLSIKHPSVNENFC  
NEKEGAQFSSHLINLLNPKGKPANQLLALRTFCNCFVGQAGQKLMMSQRESL  
MSHAIELKSGSNKNIHIALATLALNYSVCFHKDHNIEGKAQCLSLISTILEVVQD  
LEATFRLVALGTLISDDSNVQLAKSLGVDSQIKKYSSVSEPAKVSECCRFILN  
LL (SEQ ID NO: 551). Polynucleotides encoding these polypeptides are also

30 encompassed by the invention. These polypeptides share significant homology with phospholipase A2 activating protein which is thought to be important in signal transduction (see, e.g., Wang et al., Gene 161(2):237-241 (1995)).

This gene is expressed primarily in endothelial cells, to a less extent in placenta, endometrial stromal cells, osteosarcoma, testis tumor, muscle, and infant brain that are  
35 likely to be rich in blood vessels.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in vascular system, aberrant angiogenesis, tumor angiogenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system or tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in endothelial cells and several potential highly vascularized tissues and its homology to phospholipase A2 activating protein suggest that this gene may be involved in transducing signals for endothelial cells in angiogenesis or vasculogenesis.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 89

In specific embodiments, polypeptides of the invention comprise the sequence: YPNQDGDILRDQVLHEHIQRLSKVVTANHRALQIPEVYLREAPWPSAQSEIRTIS AYKTPRDKVQCILRMCSTIMNLLSLANEDSVPGADDFVPVLVFLIKANPPCLL STVQYISSFYASCLSGEESYWWMQFTA AVE (SEQ ID NO:552); YPNQDGDILR DQVLHEHIQRLSKVVTANHRALQIPEVYLREAPWPSAQSEIRTISA YKTPRDKVQ CILRMCSTIMNLLSLANEDSVPGADDFVPVLVFLIKANPPCLLSTVQYISSFYA SCLSGEESYWWMQFTA AVEFIKTI (SEQ ID NO:553); YPNQDGDILRDQVL (SEQ ID NO:554); EAPWPSAQSEI (SEQ ID NO:555); PVLVFLIKANP (SEQ ID NO:560); SGEESYWWMQFTA AVEFIKTI (SEQ ID NO:556); ADDFVPVLVFLIKANPP (SEQ ID NO:557); YKTPRDKVQCIL (SEQ ID NO:558); and/or GADDFVPVLVFLIK (SEQ ID NO:559). The translation product of this gene shares sequence homology with human ras inhibitor and yeast VPS9p which is thought to be important in golgi vacuole transport.

This gene is expressed primarily in T cells and melanocytes and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, dysfunction and disorders involving T cells and melanocytes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ras inhibitor indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating signal transduction; diagnosis and treatment of disorders involving T cells and melanocytes.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 90

This gene maps to chromosome 9 and therefore polypeptides of the invention can be used in linkage analysis as a marker for chromosome 9. The translation product of this gene shares sequence homology with neuronal olfactomedin-related ER localized protein which is thought to be important in influence the maintenance, growth, or differentiation of chemosensory cilia on the apical dendrites of olfactory neurons. In specific embodiments, polypeptides of the invention comprise the sequence: SARASTQPPAGQHGPC (SEQ ID NO:561); MPGRWRWQRDMHPARKLLSLL FLILMGTELTQD (SEQ ID NO:562); SAAPDSLLRSSKGSTRGSL (SEQ ID NO:563); AAIVWRGKSESRIAKTPGI (SEQ ID NO:564); FRGGGTLVLPPTHT PEWLIL (SEQ ID NO:567); PLGITLPLGAPETGGGD (SEQ ID NO:565); and/or CAAETWKGSQRAGQLCALLA (SEQ ID NO:566).

This gene is expressed in pineal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and endocrinological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neurological or endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 323 as residues: Leu-20 to Ala-26, Arg-32 to Arg-39, Thr-104 to Gly-112.

The tissue distribution and homology to olfactomedin-related protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for maintenance, growth, or differentiation of neuron cells in pineal gland, therefore, may be useful for diagnosis and treatment of neurological disorders in pineal gland.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 91

This gene is expressed primarily in prostate and apoptotic T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate disease and T cell dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detect abnormal activity in prostate and T cells or probably treatment of this abnormality.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 92

This gene is expressed primarily in prostate and to a lesser extent in smooth muscle cells, fibroblasts, and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in prostate or vascular system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate or vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain

tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating function of prostate or highly vascularized tissues, e.g. placenta.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 93

This gene is expressed primarily in embryos and fetal tissues stage human and to a lesser extent in a wide variety of other proliferative tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in embryonic development and cell proliferation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryonic tissues and proliferative cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of abnormalities in developing and proliferative cells and organs.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 94

The translation product of this gene shares sequence homology with transformation related protein which is thought to be important in transformation.

This gene is expressed primarily in female reproductive tissues, i.e., breast cancer cells, placenta, and ovary and to a lesser extent in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are



not limited to, cancer or dysfunction of reproductive tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproduction system,

5 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the

10 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 327 as residues: Ser-50 to Pro-61.

The tissue distribution and homology to transformation related protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of conditions caused by transformation, i.e. tumorigenesis in

15 reproductive organs, e.g. breast, placenta, and ovary.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 95

This gene is expressed primarily in testes, rhabdomyosarcoma, infant brain and to a lesser extent in some tumors and highly vascularized tissues.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumorigenesis, abnormal angiogenesis, and/or neurological disorders. . Similarly, polypeptides and antibodies directed to these polypeptides are useful in

25 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumor tissues or vascular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

30 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 328 as residues: Arg-46 to Trp-54, Pro-60 to Ile-69, Asn-116 to Ala-122, Arg-147 to Lys-153, Ser-158 to Glu-170, Ile-399 to

35 Ser-405, Pro-486 to Met-499, Pro-502 to Asp-508.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for a range of disease states including treatment of

tumor or vascular disorders and the treatment of neurological disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 96

This gene maps to chromosome 7 and therefore polynucleotides of the present invention can be used in linkage analysis as a marker for chromosome 7. The translation product of this gene is homologous to the *Clostridium perfringens* enterotoxin (CPE) receptor gene product and shares sequence homology with a human ORF specific to prostate and a glycoprotein specific to oligodendrocytes both of which are tissue specific proteins. (See e.g., Katahira et al., J Cell Biol. 136(6):1239-1247 (1997). PMID: 9087440; UI: 97242441.

This gene is expressed primarily in pancreas tumor and ulcerative colitis and to a lesser extent in several tumors and normal tissues.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pancreatic disorder, ulcerative colitis, tumors and food poisoning. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system or tumorigenic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 329 as residues: Gly-147 to Met-152, Cys-177 to Lys-188.

30 The tissue distribution and homology to prostate and oligodendrocyte-specific protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for marker of diagnosis or treatment of disorder in pancreas, ulcerative colitis, and tumors. Furthermore, identity to the human receptor for *Clostridium perfringens* enterotoxin indicates that the soluble portion of this receptor could be used in the treatment of food poisoning associated with *Clostridia perfringens* by blocking the activity of perfringens enterotoxin.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 97**

The translation product of this gene shares sequence homology with ATPase which is thought to be important in metabolism.

- 5 This gene is expressed primarily in testes and several hematopoietic cells and to a lesser extent in other tissues.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, leukemia and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 330 as residues: Leu-37 to Ala-42.

The tissue distribution and homology to ATPase indicates that polynucleotides and polypeptides corresponding to this gene are useful for marker of diagnosis and treatment of leukemia and other hematopoietic disorders.

**25 FEATURES OF PROTEIN ENCODED BY GENE NO: 98**

- In specific embodiments, polypeptides of the invention comprise the sequence: MRSARPSLGCLPSWAFSQALNI (SEQ ID NO:568); LLGLKGLAPAEISAVCE KGNFN (SEQ ID NO:569); VAHGLAWSYYIGYLRLILPELQARIR (SEQ ID NO:570); TYNQHYNLLRGAVSQRC (SEQ ID NO:571); ILLPLDCGVDPNLSM ADPNIRFLDKLPQQTGDRAGIKDRVYSN (SEQ ID NO:572); SIYELLENGQRAGT CVLEYATPLQTLFAMSQYSQAGFSGEDRLEQ (SEQ ID NO:573); AKLFCRTLE DILADAPESQNNCRLLIAYQEPADDSSFSLSQEVLRHLRQEEKEEVTGSLKTSV PSTSTMSQEPPELLISGMEKPLPLRTDFS (SEQ ID NO:574); and/or LLGLKGLA PAEISAVCEKGNFNVAHGLAWSYYIGYLRLILPEL (SEQ ID NO:575).
- 35 Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in prostate BPH and to a lesser extent in bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, benign prostatic hypertrophy or prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male urinary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 331 as residues: Ile-60 to Asn-69, Leu-106 to Asp-112, Glu-130 to Gly-136, Phe-160 to Glu-167, Pro-184 to Cys-190, Glu-197 to Ser-202, Arg-215 to Glu-221, Thr-237 to Pro-242.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of benign prostatic hypertrophy or prostate cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 99

This gene is expressed primarily in salivary gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders or injuries of the salivary gland. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of glandular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders of, or injuries to the salivary gland or other glandular tissue.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 100**

This gene maps to chromosome 15, accordingly, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 15. The translation product of this gene shares sequence homology with a *C.elegans* gene of unknown function. In specific embodiments, polypeptides of the invention comprise the sequence: DPRVRLNSLTCKHIFISLTQ (SEQ ID NO:583); TMKLLKLRRNIV KLSLYRHFTN (SEQ ID NO:576); TLILAVAASIVFIWTTMKFRI (SEQ ID NO:577); VTCQSDWRELWVDDAIVRLLFSMILFVI (SEQ ID NO:578); MVLWR PSANNQRFAFSPLSEEEEEDEQ (SEQ ID NO:580); KEPMLKESFEGMKMRS TKQEPNGNSKVNKAQEDDL (SEQ ID NO:584); and/or KWVEENVPSVTDVALP ALLDSDEERMITHFERSKME (SEQ ID NO:582). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in thyroid and to a lesser extent in osteoclastoma, kidney medulla, and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, thyroid dysfunction or cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 333 as residues: Lys-107 to Leu-124, Glu-150 to Thr-159, Pro-173 to Asp-179, Ser-192 to Ser-201.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of thyroid dysfunction or cancer.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 101**

This gene maps to chromosome 16, therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 16. In specific embodiments, polypeptides of the invention comprise the sequence:

IRHELTVLRDTRPACA (SEQ ID NO:585); and/or MDFXMALIYD (SEQ ID NO:586). Polynucleotides encoding these polypeptides are also encompassed by the invention.

5 This gene is expressed primarily in kidney cortex and to a lesser extent in adult brain, corpus colosum, hippocampus, and frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to  
10 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial  
15 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of neurological  
20 disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 102

In specific embodiments, polypeptides of the invention comprise the sequence:  
MQEMMRNQDRALSNLESIPGGYNA (SEQ ID NO:587); LRRMYTDIQEPMLSA  
25 AQEQF GGNPF (SEQ ID NO:588); ASLVSNTSSGEGSQPSRTENRDPLPNPWAP-  
QT (SEQ ID NO:589); SQSSASSGTASTVGGTTGSTASGTSGQSTTAPNLVPGV  
GASMFNTPG MQSLLQQITENPQLMQNMLSAPY (SEQ ID NO:590);  
MRSMMQSLSQNPDLAAQMMLNNPLFAGNPQLQEQMRQQLPTFLQQ (SEQ ID  
NO:591); MQNPDTLSAMSNPRAMQALLQIQQLQTLATEAPGLIPGFTPGLG  
30 ALGSTGGSSGTNGSNATPSENTSPTAGT (SEQ ID NO:592); TEPGHQQFI  
QQMLQALAGVNPQLQNPEVRFQQQLEQLSAMGFLNREANLQALATCGDINAA  
IERLLGSQPS (SEQ ID NO:593); RNPAMMQEMMRNQDRALSNLESIPGGY  
NALRRMYTDIQEPMLSAA (SEQ ID NO:594); GNPFA SLVSNTSS (SEQ ID  
NO:595); ENRDPLPNPWA (SEQ ID NO:595); GKILKDQDTLSQHGHIHD (SEQ ID  
35 NO:597); GLTVHLVIKTQNRP (SEQ ID NO:598); SELQSQMQRQLLSNPEMM  
(SEQ ID NO:599); PEISHMLNNPDIMR (SEQ ID NO:600); and/or  
RQLIMANPQMQLIQRNP (SEQ ID NO:601). Polynucleotides encoding these

polypeptides are also encompassed by the invention.

This gene is expressed primarily in breast.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of tumor systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of some types of breast cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 103

The translation product of this gene shares sequence homology with secreted serine proteases and lysozyme C precursor, which is thought to be important in bacteriolytic function. In specific embodiments, polypeptides of the invention comprise the sequence: NLCHVDCQDLLNP NLLAGIHCAKRIVS (SEQ ID NO:602); LDGFEGYSLSDWLCLAFVESKFN (SEQ ID NO:603); NENADGSFDYGLFQINSHYWCN (SEQ ID NO:604); and/or NLCHVDCQDLLNP NLLAGIHCAKRIVS (SEQ ID NO:605). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, infection. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 336 as residues: Ile-62 to Phe-70, Asn-  
5 78 to Asn-84.

The tissue distribution and homology to lysozyme C precursor indicates that polynucleotides and polypeptides corresponding to this gene are useful for boosting the monocyte-macrophage system and enhance the activity of immunoagents.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 104

This gene is expressed primarily in apoptotic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are  
15 not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,  
20 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
25 corresponding to this gene are useful for treatment and diagnosis of some immune disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 105

The translation product of this gene shares sequence homology with ARI  
30 protein of Drosophila (accession 2058299; EMBL: locus DMARIADNE, accession X98309), which is thought to be important in axonal path-finding in the central nervous system. In specific embodiments, polypeptides of the invention comprise the sequence IREVNEVIQN PAT (SEQ ID NO:606); ITRILLSHFNWDKEKLMERYF DGNLEKLFA (SEQ ID NO:607); NTRSSAQDMPCQICYNYPNSYF (SEQ ID  
35 NO:608); TGLECGHKFCMQCWSEYLTTKIMEEGMGQTISCPAHG (SEQ ID NO:614); CDILVDDNTVMRLITDSKVKLKYQHLITNSFVECNRLCLKWCPAPD CHHVVKVQYPDAKPV (SEQ ID NO:609); CDILVDDNTVMRLITDSK



VKLKYQHLITNSFVECNRLWKWCPAPDCHHVVKV (SEQ ID NO:610);  
 GCNHMVCRNQNCCKAEFCWVCLGPWEPHGSAWYNCNRYNEDDAKAARDAQE  
 RSRAALQRYL (SEQ ID NO:611); FYCNRYMNHMQSLRFEHKL YAQVKQ  
 KMEEMQQHNMSWIEVQFLKKAVDVLCQCRATLMYT (SEQ ID NO: 612);  
 5 YVFAFYLLKNNQSIIFENNQADLENATEVLSGYLERDISQDSLQDIKQKVQDKY  
 RYCESR (SEQ ID NO:613) Polynucleotides encoding these polypeptides are also  
 encompassed by the invention.

This gene is expressed primarily in adult brain, and to a lesser extent in  
 endometrial tumor, melanocytes, and infant brain.

10 Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions, which include, but are  
 not limited to, diseases or injuries involving axonal path development. Similarly,  
 polypeptides and antibodies directed to these polypeptides are useful in providing  
 15 immunological probes for differential identification of the tissue(s) or cell type(s). For  
 a number of disorders of the above tissues or cells, particularly of the central nervous  
 system, expression of this gene at significantly higher or lower levels may be routinely  
 detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,  
 serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample  
 20 taken from an individual having such a disorder, relative to the standard gene  
 expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
 individual not having the disorder.

The tissue distribution and homology to ARI protein indicates that  
 polynucleotides and polypeptides corresponding to this gene are useful for treatment of  
 25 disease states or injuries involving axonal path development, including  
 neurodegenerative diseases and nerve injury.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 106

The translation product of this gene shares sequence homology with cytochrome  
 30 b561 [Sus scrofa] which is thought to be an integral membrane protein of  
 neuroendocrine storage vesicles of neurotransmitters and peptide hormones.

This gene is expressed primarily in frontal cortex and to a lesser extent in  
 rhabdomyosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 35 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions, which include, but are  
 not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 339 as residues: Ser-18 to Pro-24.

The tissue distribution and homology to cytochrome b561 [*Sus scrofa*] indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of neurological disorders. This gene may also be important in regulation of some types of cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 107

In specific embodiments, polypeptides of the invention comprise the sequence: MWGYLFVDAAWNFLGCLICGW (SEQ ID NO:615); MHFISSGNVSAIRSSILLRXSLSYLGNCRLRVSAIFVYFLLFLLLS (SEQ ID NO:616); and/or MDQALRGSPSEGFGSTDPSPPQVGRQIPSPFPWRRLVLPKASGCFLEREWLVCVFKLRTRPGAHAHAYNSSILGGRGKGIT (SEQ ID NO:617). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in pancreas tumor and to a lesser extent in cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pancreatic tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred

epitopes include those comprising a sequence shown in SEQ ID NO: 340 as residues: Pro-22 to Phe-33.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of pancreatic tumors.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 108

This gene maps to chromosome 17 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 17. In specific embodiments, polypeptides of the invention comprise the sequence:

10 MLPALASCCHFSPPEQAARLKKLQEQEKQQKVEFRKRMEKEVSDFIQDSGQIK  
KKFQPMNKIERSILHDVVEVAGLTSFSFGEDDDCRYVMIFKKEFAPSDEELDSY  
RRGEEWDPQKAEEKRNXXKELAQRR (SEQ ID NO:618); EEEAAQQGPVVV  
SPASDYKDKYSHLIGKGAAKDAAHMLQANKTYGCXPVANKRDTRSIEEAMNE  
IRAKKRLRQSGE (SEQ ID NO:619); PPRPAQLPLTPGAGQGAGRDKAAAIRA  
15 HPGAPPLNHLLP (SEQ IDNO:620); AVPQAGGKQVFDLSPLELGYVRGMCVCV  
(SEQ ID NO:621) and/or MLPALASCCHFSPPEQAARLKKLQEQEKQQKVEFRK  
RMEKEVSDFIQDSGQIKKKFQPMNKIERSILHDVVEVAGLTSFSFGEDDDCRYV  
MIFKKEFAPSDEELDSYRRGEEWDPQKAEEKRNXXKELAQRRQEEAAQQGPVVV  
SPASDYKDKYSHLIGKGAAKDAAHMLQANKTYGCXPVANKRDTRSIEEAMNE  
20 IRAKKRLRQSGE (SEQ ID NO:622). Polynucleotides encoding these polypeptides  
are also encompassed by the invention. The translation product of this gene shares  
sequence homology with FSA-1 which may play a role as a structural protein  
component of the acrosome.

This gene is expressed primarily in fetal kidney and sperm.

25 Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
not limited to, male reproductive disorders, especially involving acrosomal dysfunction.  
Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
30 providing immunological probes for differential identification of the tissue(s) or cell  
type(s). For a number of disorders of the above tissues or cells, particularly of the male  
reproductive system, expression of this gene at significantly higher or lower levels may  
be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily  
fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or  
35 cell sample taken from an individual having such a disorder, relative to the standard  
gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an

individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 341 as residues: Glu-8 to Asn-35.

The tissue distribution and homology to FSA-1 indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of infertility due to acrosomal disfunction of sperm.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 109

This gene is expressed primarily in pituitary and to a lesser extent in epididymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 342 as residues: Met-1 to Trp-6.

Because the gene is found in both pituitary and epididymus, this indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of male reproductive disorders. This may involve a secreted peptide produced in the pituitary targeting the epididymus.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 110

In specific embodiments, polypeptides of the invention comprise the sequence: LLCPVLNSGXSWNFPHPQSQPEYSFHGFHSTRLWI (SEQ ID NO:623); and/or PSTPWFLFLLGLTCPFSTSHPRWDSIPP (SEQ ID NO:624). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in resting T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, T-cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of certain immune disorders, especially those involving T-cells.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 111

This gene is expressed primarily in cerebellum and whole brain and to a lesser extent in infant brain and fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 344 as residues: Asp-48 to Gly-55.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 112

The translation product of this gene shares sequence homology with yeast mitochondrial ribosomal protein homologous to ribosomal protein s15 of E.coli which

is thought to be important in the early assembly of ribosomes (See Accession No. M38016). This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in developmental tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development of cancers and tumors in addition to healing wounds. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and developmental expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ribosomal protein s15 of *E. coli* indicates that polynucleotides and polypeptides corresponding to this gene are useful for diseases related to the assembly of ribosomes in the mitochondria which is important in the translation of RNA into protein. Therefore, this indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of multiple tumors as well as in healing wounds which are thought to be under similar regulation as developmental tissues. Protein, as well as, antibodies directed against the protein have utility as tumor markers, in addition to immunotherapy targets, for the above listed tumors and tissues.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 113

The translation product of this gene shares sequence homology with human poliovirus receptor precursors which are thought to be important in viral binding and uptake. Preferred polypeptide fragments comprise the following amino acid sequence:  
ELISISNVALADEGEYTCSTFTMPVRTAKSLVTVLGIPQKPIITGYKSSLREKDT  
ATLNCQSSGSKPAARLTWRKGDQELHGEPTRIQEDPNGKTFTVSSSVTFQVTR  
EDDGASIVCSVNHESLKGADRSTSQRIEVLYTPTAMIRPDPPHPREGQKLLHLC  
EGRGNPVPQQYLWEKEGSVPLKMTQESALIFPFLNKSDSGTYGCTATSNMG  
YKAYYTLNVND (SEQ ID NO:625). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. gnllPID1002627).

This gene is expressed almost exclusively in human brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, susceptibility to viral disease and diseases of the CNS especially cancers of that system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 346 as residues: Leu-26 to Asp-37, Lys-53 to Ser-59.

The tissue distribution and homology to poliovirus receptor precursors indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and prevention of diseases that involve the binding and uptake of virus particles for infection. It might also be helpful in genetic therapy where the goal is to insert foreign DNA into infected cells. With the help of this protein, the binding and uptake of this foreign DNA might be aided. In addition, it is expected that over expression of this gene will indicate abnormalities involving the CNS, particularly cancers of that system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 114

The translation product of this gene shares sequence homology with YO87\_CAEEL hypothetical 28.5 KD protein ZK1236.7 in chromosome III of *Caenorhabditis elegans* in addition to alpha-1 collagen type III (See Accession No. gil537432). One embodiment for this gene is the polypeptide fragment(s) comprising the following amino acid sequence: VPELPDRVHQLHQA VQGICALGRPGFPGGPTH SGHHKSHPGPAGGDYNRCDRPGQVHLHNPRGTGRRGQLHPTAGPGVHRRACPSQQLPHRLGPGVPCPSPLTPVLPSWTQSWCG LPGYTSSS (SEQ ID NO:630). An additional embodiment is the polynucleotide fragment(s) encoding these polypeptide fragments

This gene is expressed primarily in brain cells and to a lesser extent in activated B and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegeneration and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 347 as residues: Glu-34 to Glu-39, Gly-51 to Ser-72, Ala-88 to Glu-93, Gln-100 to Val-105.

The tissue distribution and homology to YO87\_CAEEL hypothetical 28.5 KD protein ZK1236.7 in chromosome III of *Caenorhabditis elegans* as well as to a conserved alpha-1 collagen type III protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons' Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorders. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 115

The translation product of this gene shares sequence homology with alpha 3 type IX collagen which is thought to be important in hyaline cartilage formation via its ability to uptake inorganic sulfate by cells (See Accession No. gil975657). One embodiment of this gene is the polypeptide fragment comprising the following amino acid sequence: SLRRPRSAAXQTLTTLSSVSSASSSALPGSREPCDPRAPPPR SGSAASCCSCCCSCPRRRAPLRSPRGSKRRIRQREVVDLYNGMCLQGPAVPG RDGSPGANGIPGTPGIPGRDGFKEGKECLRESFEESWTPNYKQCSWSSLNY GIDLGKIAECTFTKMRSNSALRVLFSGSLRLKCRNACCQRWYFTFNGAECSSGP LPIEAIYLDQGSPEMNSTINIHRTSSVEGLCEGIGAGLVDVAIWVGTCSDYPKG DASTGWNSVSRIIIIEELPK (SEQ ID NO:634). An additional embodiment are the



polynucleotide fragments encoding this polypeptide fragment.

This gene is expressed primarily in smooth muscle and to a lesser extent in synovial tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias i.e., spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid and autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to alpha 3 type IX collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of diseases associated with the mutation in this gene which leads to the many different types of chondrodysplasias. By the use of this product, the abnormal growth and development of bones of the limbs and spine could be routinely detected or treated in utero since the protein or muteins thereof could affect epithelial cells early in development and later the chondrocytes of the developing craniofacial structure.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 116

The translation product of this gene shares sequence homology with retrovirus-related reverse transcriptase which is thought to be important in viral replication. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: TKKENC RPASLMNIDTKILNKILMNQ (SEQ ID NO:640). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments (See Accession No. pirlA25313IGNHUL1).

This gene is expressed primarily in human meningioma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, retroviral diseases such as AIDS, and possibly certain cancers due to transactivation of latent cell division genes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to retrovirus-related reverse transcriptase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of diseases and maladies associated with retroviral infection since a functional reverse transcriptase (RT) or RT-like molecule is an integral component of the retroviral life cycle.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 117

The translation product of this gene shares sequence homology with an unknown gene from *C. elegans*, as well as weak homolog with mammalian metaxin, a gene contiguous to both thrombospondin 3 and glucocerebrosidase, is known to be required for embryonic development. Preferred polypeptide fragments comprise the following amino acid sequence: MCNLPKVVCRANAEYMSPSGKVPXXHVGNGQ VVSELGPIVQFVKAKGHSLSGLEEVQKAEMKAYMELVNNMLLTAELYLQWC DEATVGXITHXRYGSPYPWPLXHLAYQKQWEVKRXXKAIGWGKKTLTDQVLE DVDQCCQALSQRLGTQPYFFNKQPTELDALVFGHLYTILTTQLTNDELSEKVKN YSNLLAFCRRI EQHYFEDRGKGRLS (SEQ ID NO:641); MCNLPKVVCRANAE YMSPSGKVPXXHVGNGQVVSELGPIVQFVK (SEQ ID NO:642). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. gill326108).

This gene is expressed primarily in fetal tissues and to a lesser extent in hematopoietic cells and tissues, including spleen, monocytes, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer; lymphoproliferative disorders; inflammation; chondrosarcoma, and Gaucher disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification

of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and embryonic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression in embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation or cellular division. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and survival of hematopoietic cell lineages. Thus, this gene may be useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 118

The translation product of this gene shares sequence homology with reverse transcriptase which is important in the synthesis of a cDNA chain from an RNA molecule, and is a method whereby the infecting RNA chains of retroviruses are transcribed into their DNA complements. One embodiment for this gene is the polypeptide fragment comprising the following amino acid sequence:  
MXXXNSHITIFTLNVNGLNAPNERHRLANWISQDQVCCIQETHLTGRDTHRL  
KIKGWRKIYQANGKQKK (SEQ ID NO:647). An additional embodiment is the polynucleotide fragments comprising polynucleotides encoding these polypeptide fragments (See Accession No. gil2072964).

This gene is expressed primarily in skin and to a lesser extent in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, hematopoietic disorders; inflammation; disorders of immune surveillance. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the epidermis and/or hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and

wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5       The tissue distribution and homology to reverse transcriptase indicates that polynucleotides and polypeptides corresponding to this gene are useful for cancer therapy. Expression in the skin also indicates that this gene is useful in wound healing and fibrosis. Expression by neutrophils also indicates that this gene product plays a role in inflammation and the control of immune surveillance (i.e. recognition of viral  
10       pathogens). Reverse transcriptase family members are also useful in the detection and treatment of AIDS.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 119

15       The translation product of this gene shares sequence homology with reverse transcriptase which is important in the synthesis of a cDNA copy of an RNA molecule, and is a method whereby a retrovirus reverse-transcribes its genome into an inheritable DNA copy.

      This gene is expressed primarily in the frontal cortex of brain.

20       Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders  
25       of the above tissues or cells, particularly of the CNS and peripheral nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,  
30       the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

      The tissue distribution and homology to reverse transcriptase suggest that this is useful in the treatment of cancer and AIDS. The expression in brain indicates that it plays a role in neurodegenerative disorders and in neural degeneration.

35

**FEATURES OF PROTEIN ENCODED BY GENE NO: 120**

One embodiment of this gene has homology to a hypothetical protein in *Schizosaccharomyces pombe* (See Accession No. 2281980). Another embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

5 IYHLHSWIFFHFKRAFCMCFITMKVIHAHCSKLRKCXNAQISVFCTTLTASYPT  
(SEQ ID NO:651). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. This gene maps to chromosome 18, and therefore, may be used as a marker in linkage analysis for chromosome 18.

10 This gene is expressed primarily in adult hypothalamus and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disorders; endocrine function; and vertigo. Similarly,  
15 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, CNS and peripheral nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded  
20 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
25 corresponding to this gene are useful for the treatment and diagnosis of neurodegenerative disorders; diagnosis of tumors of a brain or neuronal origin; treatments involving hormonal control of the entire body and of homeostasis, behavioral disorders, such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and  
30 panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 121**

35 The translation product of this gene shares sequence homology with the human IRLB protein which is thought to be important in binding to a c-myc promoter element and thus regulating its transcription (See Accession No. gi33969). This gene maps to

chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in brain and breast and to a lesser extent in a variety of hematopoietic tissues and cells.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer of the brain and breast; lymphoproliferative disorders; neurodegenerative diseases. Similarly, polypeptides and antibodies directed to these  
10 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS, breast, and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial  
15 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of cancer of the  
20 brain, breast, and hematopoietic system. In addition, it may be useful for the treatment of neurodegenerative disorders, as well as disorders of the hematopoietic system, including defects in immune competency and inflammation. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

25

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 122

The translation product of this gene shares sequence homology with an ATP synthase, a key component of the proton channel that is thought to be important in the translocation of protons across the membrane.

30 This gene is expressed primarily in T-cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, T cell lymphoma. Similarly, polypeptides and antibodies directed to these  
35 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or

lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ATP synthase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of defects in proton transport, homeostasis, and metabolism, as well as the diagnosis and treatment of lymphoma. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 123

This gene maps to chromosome 15, and therefore, may be used as a marker in linkage analysis for chromosome 15.

This gene is expressed primarily in a variety of fetal tissues, including fetal liver, lung, and spleen, and to a lesser extent in a variety of blood cells, including eosinophils and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer (abnormal cell proliferation); T cell lymphomas; and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetus and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions involving cell proliferation. Expression of this gene in fetal tissues, as well as in a variety of blood cell lineages indicates that it may play a role in either cellular proliferation; apoptosis; or cell survival. Thus it may be useful in the management and

treatment of a variety of cancers and malignancies. In addition, its expression in blood cells suggest that it may play additional roles in hematopoietic disorders and conditions, and could be useful in treating diseases involving autoimmunity, immune modulation, immune surveillance, and inflammation..

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 124

This gene is expressed primarily in placenta and to a lesser extent in pineal gland and rhabdomyosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental, endocrine, and female reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For  
15 a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely, e.g., immune], expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having  
20 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 357 as residues: Leu-69 to Val-76.

The tissue distribution indicates that polynucleotides and polypeptides  
25 corresponding to this gene are useful for diagnosis and treatment of disorders in development. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 125

This gene is expressed primarily in benign prostatic hyperplasia.

Therefore, polynucleotides and polypeptides of the invention are useful as  
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of benign prostatic hyperplasia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive



system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of benign prostatic hyperplasia. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 126

This gene is expressed primarily in apoptotic T-cells and to a lesser extent in suppressor T cells and ulcerative colitis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases involving premature apoptosis, and immunological and gastrointestinal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders involving inappropriate levels of apoptosis, especially in immune cell lineages. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases (such as AIDS), and leukemia.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 127**

This gene is expressed primarily in Raji cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and T cell autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 360 as residues: Asp-23 to Gly-29.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammation and T cell autoimmune disorders. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases (such as AIDS), and leukemia.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 128**

The translation product of this gene shares sequence homology with an *C. elegans* coding region C47D12.2 of unknown function (See Accession No. gnllPIDe348986). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: EDDGFNRSIHEVILKNITWY SERVLTEISLGSLLLLVVIRTIQYNMTRTRDKYLHTNCLAALANMSAQFRSLHQY AAQRIISLFSLLSKKHNVLEQATQSLRGSLSNDVPLPDYAQDLNVIEEVIRMM LEIINSCLTNSLHHNPNLVALLYKRDLFEQFRTHPSFQDIMQNIDLVISFFSSRLQAGS (SEQ ID NO:657); EDDGFNRSIHEVILKNITWYSERVLTEISLGSLLLLVV (SEQ ID NO:658); RTIQYNMTRTRDKYLHTNCLAALANMSAQFRSLHQYAAQ RIISLFSLLSKKHNVLEQATQSLRGSLSNDVPLPDYAQD (SEQ ID NO:661); SCLTNSLHHNPNLVALLYKRDLFEQFRTHPSFQDIMQNIDLVISFFSSRLQAGS (SEQ ID NO:660). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. This gene maps to

chromosome 18, and therefore, may be used as a marker in linkage analysis for chromosome 18.

This gene is expressed primarily in smooth muscle and to a lesser extent in fetal liver.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, atherosclerosis and other cardiovascular and hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing  
10 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample  
15 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of circulatory system  
20 disorders such as atherosclerosis, hypertension, and thrombosis. In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the  
25 expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 129

30 The translation product of this gene shares sequence homology with a ribosomal protein which is thought to be important in cellular metabolism, in addition to the *C.elegans* protein F40F11.1 which does not have a known function at the current time (See Accession No. gnllPIDe244552). Preferred polypeptide fragments comprise the following amino acid sequence:

35 MADIQTERA YQKQPTIFQNKKRVLGGETGKEKLPRVTNKNIGLGFKDT  
PRLLRGTYIDKKCPFTGNVSIRGRILSGVVTQDEDAEDHCHPPRLSALHPQVQ  
PLREAPQEHVCTPVPL LQGRPDR (SEQ ID NO:662): MKMQRTTVIRRDYLH

YIRKYNRFEKRRHKNMSVHLSPCFRDVQIGDIVTVGECRPLSKTVRFNVLKVTK  
 AAGTKKQFQKF (SEQ ID NO:663); MADIQTERAYQKQPTIFQNKKRVLGET  
 GK (SEQ ID NO:664); HCHPPRLSALHPQVQPLREAPQEHVCTPVPL LQGRPDR  
 (SEQ ID NO:666); NIGLGFKDTPRLLRGTYIDKKCPFTGNVSIRGRILSGVVTQ  
 5 (SEQ ID NO:669); MKMQRTIVIRRDYLHYIRKYNRFEKRRHKNMSVHLSP (SEQ  
 ID NO:667); CFRDVQIGDIVTVGECRPLSKTVRFNVLKVTKAAGTKKQFQKF  
 (SEQ ID NO:668). Also preferred are polynucleotide fragments encoding these  
 polypeptide fragments.

10 This gene is expressed primarily in Wilm's tumor and to a lesser extent in  
 thymus and stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions, which include, but are  
 not limited to, diseases affecting RNA translation. Similarly, polypeptides and  
 15 antibodies directed to these polypeptides are useful in providing immunological probes  
 for differential identification of the tissue(s) or cell type(s). For a number of disorders  
 of the above tissues or cells, particularly of the Wilm's tumors, expression of this gene  
 at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,  
 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial  
 20 fluid or spinal fluid) or another tissue or cell sample taken from an individual having  
 such a disorder, relative to the standard gene expression level, i.e., the expression level  
 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred  
 epitopes include those comprising a sequence shown in SEQ ID NO: 362 as residues:  
 Thr-11 to Asp-20.

25 The tissue distribution and homology to a ribosomal protein indicates that  
 polynucleotides and polypeptides corresponding to this gene are useful for diseases  
 affecting RNA translation.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 130

30 The translation product of this gene shares sequence homology with a yeast  
 DNA helicase which is thought to be important in global transcriptional regulation (See  
 Accession No. gnllPIDe243594). One embodiment for this gene is the polypeptide  
 fragments comprising the following amino acid sequence: IFYDSDWNPTVDQQA  
 MDRAHRLGQTKQVTYRRLICKGTIEERILQRAKEKSEIQRMVISG (SEQ ID  
 35 NO:670); TRMIDLLEEYMVYRKHTYXRLDGSSKISERRDMVADFQNRNDI  
 FVFLSTRAGGLGINLTAXDTVHF (SEQ ID NO:671); TRMIDLLEEYMVYRK  
 HTYXRLDGSSKISERRDM (SEQ ID NO:674); RRDMMVADFQNRNDIFVFL

STRAGGLGINLTAXDTVHF (SEQ ID NO:675), IFYDSDWNPTVDQQAMD  
RAHRLGQTKQVTVYRLICKG (SEQ ID NO:676); RLICKGTIEERILQRAK  
EKSEIQRMVISG (SEQ ID NO:678). An additional embodiment is the polynucleotide  
fragments encoding these polypeptide fragments.

5 This gene is expressed primarily in amygdala.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
not limited to, diseases and disorders of the brain. Similarly, polypeptides and  
10 antibodies directed to these polypeptides are useful in providing immunological probes  
for differential identification of the tissue(s) or cell type(s). For a number of disorders  
of the above tissues or cells, particularly of the central nervous system, expression of  
this gene at significantly higher or lower levels may be routinely detected in certain  
tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,  
15 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an  
individual having such a disorder, relative to the standard gene expression level, i.e.,  
the expression level in healthy tissue or bodily fluid from an individual not having the  
disorder.

The tissue distribution and homology to a DNA helicase indicates that  
20 polynucleotides and polypeptides corresponding to this gene are useful for diseases  
affecting RNA transcription, particularly developmental disorders and healing wounds  
since the later are though to approximate developmental transcriptional regulation.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 131

25 This gene is expressed primarily in prostate and to a lesser extent in amygdala  
and pancreatic tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
30 not limited to, prostate enlargement and gastrointestinal disorders, particularly of the  
pancreas and gall bladder. Similarly, polypeptides and antibodies directed to these  
polypeptides are useful in providing immunological probes for differential identification  
of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,  
particularly of the reproductive system, expression of this gene at significantly higher or  
35 lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded  
tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or  
another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of prostate diseases, including benign prostatic hyperplasia and prostate cancer. In addition, the tissue distribution in tumors of the pancreas indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tissues where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 132

This gene is expressed primarily in adult lung and to a lesser extent in hypothalamus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pulmonary diseases and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary and respiratory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of pulmonary and respiratory disorders such as emphysema, pneumonia, and pulmonary edema and emboli. In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental

disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 133

5 This gene is expressed primarily in human liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cirrhosis of the liver and other hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver disorders such as cirrhosis, jaundice, and Hepatitis. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 134

This gene is expressed primarily in fetal kidney and to a lesser extent in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development and regeneration of liver and kidney and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and excretory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level; i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 367 as residues: Pro-70 to Arg-77, Tyr-102 to Thr-107.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the kidney and liver; such as cirrhosis, kidney failure, kidney stones, and liver failure, hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells. In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 135

This gene is expressed primarily in brain, bone marrow, and to a lesser extent in placenta, T cell, testis and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative and immunological diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 368 as residues: Met-1 to His-6.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also



play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, or sexually-linked disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 136

- 5 Translation product of this gene is homologous to the human WD repeat protein HAN11. Preferred polypeptide fragments comprise the following amino acid sequence:
- MSLHGKRKEIYKYEAPWTVYAMNWSVRPDKRFRALGVSFVEEYNNKVQLVG  
 LDEESSEFICRNTFDHPYPTTKLMWIPDTKGVYPDLLATSGDYLRVWRVGETET  
 10 RLECLLNNNKNSDFCAPLTSFDWNEVDPYLLGTSSIDTTCTIWGLETGQVLGRV  
 NLVSGHVKTQLIAHDKEVYDIAFSRAGGGRDMFASVGADGSVRMFDLRHLEH  
 STIYEDPQHHPDLLRLCWNKQDPNYLATMAMDGMEVVILDVRVPAHLXPGTTIE  
 HVSMALLGPHIHPATSALQRMTRLSSTSSKCPEPLRTLSPWPTQLXGEINNVQ  
 WASTQPELSPSATTTAWRYSECSVGGAVPTRQGLLYFLPLPHPQS (SEQ ID  
 15 NO:679); MSLHGKRKEIYKYEAPWTVYAMNWSVRPDKRFRALGVSFV  
 EEYNNKVQLVGLDEESSEFICRNTFDHPYPTTKLMWIPDTKGVYPDLLATSGDY  
 LRVWRVGETETRLECLLNNNKNSDFCAPLTSFDWNEVDPYLL (SEQ ID  
 NO:680); SFDWNEVDPYLLGTSSIDTTCTIWGLETGQVLGRVNLVSGHVK  
 TQLIAHDKEVYDIAFSRAGGGRDMFASVGADGSVRMFDLRHLEHSTIYEDPQH  
 20 HPLLRLCWNKQDPNYLATMAMDGMEVVILDVRVPAHLXPGTTI (SEQ ID  
 NO:681); VGADGSVRMFDLRHLEHSTIYEDPQHHPDLLRLCWNKQDPNYLA  
 TMAMDGMEVVILDVRVPAHLXPGTTIEHVSMALLGPHIHPATSALQRMTRLS  
 SGTSSKCPEPLRTLSPWPTQLXGEINNVQWASTQPELSPSATTTAWRYSECSV  
 GAVPTRQGLLYFLPLPHPQS (SEQ ID NO:682). Also preferred are polynucleotide  
 25 fragments encoding these polypeptide fragments.

This gene is expressed primarily in placenta, embryo, T cell and fetal lung and to a lesser extent in endothelial, tonsil and bone marrow.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
 30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological and developmental diseases in addition to cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the  
 35 immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or

cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 369 as residues: Gly-19 to Gln-28, Pro-36 to Phe-42.

5       The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above  
10       listed tumors and tissues. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

#### 15       **FEATURES OF PROTEIN ENCODED BY GENE NO: 137**

      This gene is expressed primarily in TNF and INF induced epithelial cells, T cells and kidney.

      Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
20       biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory conditions particularly inflammatory reactions in the kidney. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of renal  
25       system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
30       individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 370 as residues: Thr-67 to Gly-72, Gln-132 to Ala-145, Arg-150 to Pro-157.

      The tissue distribution indicates that the protein products of this gene are useful for treating the damage caused by inflammation of the kidney.

35

**FEATURES OF PROTEIN ENCODED BY GENE NO: 138**

This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. D63485).

5 This gene is expressed primarily in breast cancer and colon cancer and to a lesser extent in thymus and fetal spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers, especially of the breast and colon tissues. Similarly,  
10 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,  
15 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon and breast origins indicates that  
20 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors; in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

**25 FEATURES OF PROTEIN ENCODED BY GENE NO: 139**

This gene maps to chromosome 17, and therefore, can be used as a marker for linkage analysis from chromosome 17.

This gene is expressed primarily in CD34 positive cells, and to lesser extent in activated T-cells and neutrophils.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunologically related diseases and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing  
35 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and hematopoietic system, expression of this gene at significantly higher or lower levels

may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in CD34, T-cell and neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of hematopoietic disorders and immunologically related diseases, such as anemia, leukemia, inflammation, infection, allergy, immunodeficiency disorders, arthritis, asthma, immune deficiency diseases such as AIDS.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 140

This gene was recently cloned by another group, who called the gene KIAA0313 gene. (See Accession No. d1021609.) Preferred polypeptide fragments comprise the amino acid sequence:

LYATATVISSPSTEXLSQDQGDRASLDAADSGRGSWTSCSSGSHDNIQTIQ  
HQRSWETLPFGHTHFDYSGDPAGLWASSSHMDQIMFSDHSTKYNRQNSRES  
LEQAQSRASWASSTGYWGEDSEGDGTIKRRGGKDVSEAESSSLTSVTTEETK  
PVPMPAHIAVASSTTKGLIARKEGRYREPPPTPPGYIGIPITDFPEGHSHPARKP  
PDYNVALQRSRMVARSSDTAGPSSVQQPHGHPTSSRPVNKPQWHKXNESDPR  
LAPYQSQGFSTEEDEDEQVSAV (SEQ ID NO:683); HMDQIMFSDHSTKYNRQ  
NQSRESLEQAQSRASWASSTGYWGE (SEQ ID NO:684); SVTTEETKPVPMP  
AHIAVASSTTKGLIARKEGRYREPPPTPPGYIGIPITD (SEQ ID NO:685); and  
VALQRSRMVARSSDTAGPSSVQQPHGHPTSSRPVNKPQW

HKXNESDPRLAPYQSQGF (SEQ ID NO:686). Also preferred are polynucleotide fragments encoding these polypeptide fragments. This gene maps to chromosome 4, and therefore, may be used as a marker in linkage analysis for chromosome 4 (See Accession No. AB002311).

This gene is expressed primarily in ovarian cancer, tumors of the Testis, brain, and colon.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, ovarian, testicle, brain and colon cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male and female reproductive systems,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,  
5 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon, ovary, testis, and brain origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where  
10 expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 141

15 This gene is expressed primarily in spleen and colon cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, colon cancer and immunological disorders. Similarly, polypeptides and  
20 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gastrointestinal tract and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,  
25 serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon, ovary, and breast origins indicates  
30 that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

35

**FEATURES OF PROTEIN ENCODED BY GENE NO: 142**

Translation product is homologous to T cell translocation protein, a putative zinc finger factor (See Accession No. 340454), as well as to the G-protein coupled receptor TM5 consensus polypeptide (See Accession No. R50734). Preferred polypeptide

5 fragments comprise the following amino acid sequence:

CLLFVVFVSLGMRCLEFWTIVYNVLYLKHKCNVLLCYHLCSI (SEQ ID NO:687);  
ACSKLIPAFEMVMRAKDNVYHLDCAFQCLNQRXCVDKFFLKNNXXLCQT  
DYEEGLMKEGYAPXVR (SEQ ID NO:688). Also preferred are polynucleotide  
fragments encoding these polypeptide fragments.

10 This gene is expressed primarily in fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders including brain cancer. Similarly, polypeptides  
15 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Central Nervous System, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,  
20 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
25 corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with  
30 the developing embryo.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 143**

Translation product for this gene has significant homology to the Fas ligand, which is a cysteine-rich type II transmembrane protein/tumor necrosis factor receptor  
35 homolog. Mutations within this protein have been shown to result in generalized lymphoproliferative disease leading to the development of lymphadenopathy and autoimmune disease (See Medline Article No. 94185175). Preferred polypeptide

fragments comprise the following amino acid sequence:

SALSEPGAPDRRRPCPESVPRRPDDEQWPPPTALCLDVAPLPPSS (SEQ ID NO:689). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. 473565).

5 This gene is expressed primarily in osteoblasts, lung, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoblast-related, pulmonary, neurological, and immunological  
10 diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded  
15 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 376 as residues: Trp-33 to Thr-40, Lys-  
20 45 to Ile-63.

The tissue distribution in osteoblasts, lung, and brain combined with its homology to the Fas ligand indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as,  
25 antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Because the Fas ligand gene is known to be expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including asthma, immune deficiency diseases such as AIDS  
30 and leukemia, and various autoimmune disorders including lupus and arthritis.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 144

This gene shares sequence homology with a 21.5 KD transmembrane protein in the SEC15-SAP4 intergenic region of yeast. (See Accession No. 1723971.) Preferred  
35 polypeptide fragments comprise the amino acid sequence:

AHASESGERWWACCGVRFGLRSIEAIGRSCCHDGPGLVANRGRFRFKWAIEL  
SGPGGSGRGRSDRGSGQGDSLYPVGYLDKQVPDTSVQETDRILVEKRCWDIAL

GPLKQIPMNLFIMYMAGNTISIFPTMMVCMMAWRPIQALMAISATFKMLESSSQ  
 KFLQGLVYLIGNLMGLALAVYKCQSMGLLPTHASDWLAFIEPPERMEFSGG  
 GLLL (SEQ ID NO:691); PVGYLDKQVPDTSVQETDRILVEKRCWDIALGPLKQ  
 IPMNLFI (SEQ ID NO:693); and ATFKMLESSSQKFLQGLVYLIGNLMGLALAV  
 5 YKCQSMGLLPTHASD (SEQ ID NO:692). Also preferred are polynucleotide  
 fragments encoding these polypeptide fragments.

This gene is expressed primarily in osteoclastoma, hemangiopericytoma, liver,  
 lung.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 10 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions, which include, but are  
 not limited to, osteoclastoma, hemangiopericytoma, liver and lung tumors. Similarly,  
 polypeptides and antibodies directed to these polypeptides are useful in providing  
 immunological probes for differential identification of the above tissue(s) or cell  
 15 type(s). For a number of disorders of the above tissues or cells, particularly of the lung  
 and liver systems, expression of this gene at significantly higher or lower levels may be  
 routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily  
 fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or  
 cell sample taken from an individual having such a disorder, relative to the standard  
 20 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
 individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
 corresponding to this gene are useful for diagnosing osteoclastoma,  
 hemangiopericytoma, liver and lung tumors.

## 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 145

Translation product of this gene shares homology with the glucagon-69 gene  
 which may indicate this gene plays a role in regulating metabolism. (See Accession No.  
 A60318) One embodiment for this gene is the polypeptide fragments comprising the  
 30 following amino acid sequence:  
 PTTKLDIMEKKKHIIQIRFPSFYHKLVDSGRMRSKRETRREDSDTKHNL (SEQ ID  
 NO:694). An additional embodiment is the polynucleotide fragments encoding these  
 polypeptide fragments.

This gene is expressed primarily in brain, kidney, colon, and testis.

35 Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions, which include, but are



not limited to, brain, kidney, colon, and testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, neurological, circulatory, and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of brain, kidney, colon, and testis origins, indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

## 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 146

The translation product of this gene shares sequence homology with goliath protein which is thought to be important in the regulation of gene expression during development. Protein may serve as a transcription factor. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

30 TEHILAVMITELRGKDILSYLEKNISVQMTIAVGTRMPPKNFSRGS�VFVSISFIV  
 LMISSAWLIFYFIQKIRYTNARDRNQRR LGDAAKKAISKLTTRTVKKGDKETD  
 PDFDHCAVCIESYKQNDVVRILPCKHVFHKSCVDPWLSEHCTCPMCKLNILKA  
 LGIV (SEQ ID NO:695); TEHILAVMITELRGKDILSYLEKNISVQMTIAVGTRMP  
 PKNFSRGS�VFVSISFIVLM IISSAWLIFYF (SEQ ID NO:697); SISFIVLMISSA  
 35 WLIFYFIQKIRYTNARDRNQRR LGDAAKKAISKLTTRTVKKGDKE (SEQ ID  
 NO:698); VKKGDKETDPDFDHCAVCIESYKQNDVVRILPCKHVFHKSCVDP

WLSEHCTCPMCKLNILKALGIV (SEQ ID NO:699). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments (See Accession No. 157535). Moreover, another embodiment is the polynucleotide fragments encoding these polypeptide fragments:

- 5 MTHPGTEHIIA VMITELRGKDILSYLEKNISVQMTIA VGTRMPPKNFSRGS  
LVFVSISFIVLMISSAWLIFYFIQKIRYTNARDRNQRR LGDAAKKAISKLTTRTV  
KKGDKETDPDFDHCAVCIESYKQNDVVRILPCKHVFHKSCVDPWLSEHCTCP  
MCKLNILKALGIVPNLPCTDNVAFD MERLTRTQAVNRRSALGDLAGDNSLGLE  
PLRTSGISPLPQDGELTPRTGEINIA VTKEWFIIASFGLLSALTLCYMIIRATASLN  
10 ANEVEWF (SEQ ID NO:696); MTHPGTEHIIA VMITELRGKDILSYLEKNISVQM  
TIA VGTRMPPKNFSRGS LVFVSISFIVLMISSAWLIFYFIQKIRYTNARDRNQRR  
LGDAAKKAISKLTTRT (SEQ ID NO:700); AAKKAISKLTTRTVKKGDKE  
TDPDFDHCAVCIESYKQNDVVRILPCKHVFHKSCVDPWLSEHCTCPMCKLNIL  
KALGIVPNLPC (SEQ ID NO:701); TQAVNRRSALGDLAGDNSLGLEPLRTSGI  
15 SPLPQDGELTPRTGEINIA VTKEWFIIASFGLLSALTLCYMIIRATASLNANEVEW  
F (SEQ ID NO:702); PLHGVADHLGCDPQTRFFVPPNIKQWIAL LQRGNCTF  
KEKISR AAFHNAVAVVIYNNKSKEEPVTMTHPGTEHIIA VMITELRGKDILSYLE  
KNISVQMTIA VGTRMPPKNFSRGS LVFVSISFIVLMISSAWLIFYFIQKIRYTN  
RDRNQRR LGDAAKKAISKLTTRTVKKGDKETDPDFDHCAVCIESYKQNDVVR  
20 LPCKHVFHKSCVDPWLSEHCTCPMCKLNILKALGIVPNLPCTDNVAFD MERLT  
RTQAVNRRSALGDLAGDNSLGLEPLRTSGISPLPQDGELTPRTGEINIA VTKEW  
FIASFGLLSALTLCYMIIRATASLNANEVEWF (SEQ ID NO:703); and  
HGVADHLGCDPQTRFFVPPNIKQWIAL LQRGNCTFKEKISR AAFHNAVAVVIY  
NNKSKEE (SEQ ID NO:704). An additional embodiment is the polynucleotide  
25 fragments encoding these polypeptide fragments. When tested against Jurkat cell lines,  
supernatants removed from cells containing this gene activated the GAS pathway.  
Thus, it is likely that this gene activates immune cells through the JAKS/STAT signal  
transduction pathway.

- 30 This gene is expressed primarily in macrophage, breast, kidney and to a lesser  
extent in synovium, hypothalamus and rhabdomyosarcoma.

- Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
not limited to, schizophrenia and cancer. Similarly, polypeptides and antibodies directed  
35 to these polypeptides are useful in providing immunological probes for differential  
identification of the tissue(s) or cell type(s). For a number of disorders of the above  
tissues or cells, particularly of the immune and neural system, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to zinc finger protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of schizophrenia, kidney disease and other cancers. The tissue distribution in macrophage, breast, and kidney origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of tumors within these tissues, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 147

The translation product of this gene shares sequence homology with HNP36 protein, an equilibrative nucleoside transporter, which is thought to be important in gene transcription as well as serving as an important component of the nucleoside transport apparatus (See Accession No. 1845345). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MSGQGLAGFFASVAMICAIASGSELSESAFGYFITACAVIILTIICYLGLPRLEFYR  
 YYQQLKLEGPGEQETKLDLISKGEEPRAGKEESGVSVSNSQPTNESHSHKAILK  
 NISVLAFSVCFTITIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLT FNIFDWLG  
 RSLTAVFMWPGKDSRWLPSWXLARLVFVPLLLLCNIKPRRYLTVVFEHDAWFI  
 FFMAAFASNGYLASLCMCFGPKKVPAEAETAEPSPSSCVVWWHWGLFS  
 PSCSGQLCDKGWTEGLPASLPVCLLPARGDPEWSGGFFF (SEQ ID  
 NO:705); MSGQGLAGFFASVAMICAIASGSELSESAFGYFITACAVIILTIIC  
 YLGLPRLEFYRYYQQLKLE GPGEQETKLDLISKGEEPRAGKEESGVSVSNSQ  
 PTNESHSH (SEQ ID NO:706); SGVSVSNSQPTNESHSHKAILKNISVLAFSVCFI  
 FTITIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLT FNIFDWLGRS (SEQ ID  
 NO:707), TIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLT FNIFDWLGRSLTAVF  
 MWPGKDSRWLPSWXLARLVFVPLLLLCNIK PRRYLTVVFEHDA (SEQ ID  
 NO:708); FGPKKVPAEAETAEPSPSSCVVWWHWGLFSPSCSGQLCDK

GWTEGLPASLPVCLLPLPSARGDPEWSGGFFF (SEQ ID NO:709). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in eosinophils and aortic endothelium and to a lesser extent in umbilical vein endothelial cell and thymus.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential  
10 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having  
15 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to HNP36 protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of blood neoplasias and other hematopoietic disease.

## 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 148

This gene is expressed primarily in breast cancer cell lines, thymus stromal cells, and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as  
25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, endocrine and female reproductive system diseases including breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell  
30 type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard  
35 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of endocrine disorders. In addition, the tissue distribution in tumors of thymus, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 149

Translation product of this gene has homology to pmt1 and pmt 2, two conserved schizosaccharomyces pombe genes. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

DDDGFEIVPIEDPAKHRILDPEGLALGAVIASSKKAKRDLIDNSFNRYTFNEDEG  
 15 ELPEWVFVQEEKQHRIRQLPVGKKEVEHYRKRWREINARPIXXXXXXXXXXXXX  
 XXXXXXLEQTRKKAEEVVNTVDIXRTRES (SEQ ID NO:710);  
 DDDGFEIVPIEDPAKHRILDPEGLALGAVIASSKKAKRDLIDNSFNRYTF (SEQ  
 ID NO:711); KRWREINARPIXXXXXXXXXXXXXXXXXXXXXLEQTRKKAEE  
 AVVNTVDIXRTRES (SEQ ID NO:712). An additional embodiment is the  
 20 polynucleotide fragments encoding these polypeptide fragments (See Accession No.  
 e1216734).

This gene is expressed primarily in retina and ovary and to a lesser extent in breast cancer cell, epididymus and osteosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 25 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions, which include, but are  
 not limited to, neuronal growth disorders, cancer and reproductive system disorders.  
 Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
 providing immunological probes for differential identification of the tissue(s) or cell  
 30 type(s). For a number of disorders of the above tissues or cells, particularly of the  
 neural and reproductive system, expression of this gene at significantly higher or lower  
 levels may be routinely detected in certain tissues (e.g., cancerous and wounded  
 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or  
 another tissue or cell sample taken from an individual having such a disorder, relative to  
 35 the standard gene expression level, i.e., the expression level in healthy tissue or bodily  
 fluid from an individual not having the disorder. Preferred epitopes include those  
 comprising a sequence shown in SEQ ID NO: 382 as residues: Met-1 to Gly-7.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis or treatment of reproductive system disease and cancers.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 150

One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MIKDKGRARTALTSSQPAHLCPENPLLHLKAAVKEKKRNKKKKTIGSPKRIQS  
 PLNNKLLNSPAKTLPGACGSPQKLIDGFLKHEGPPAEKPLEELSASTSGVPGLS  
 10 SLQSDPAGCVRPPAPNLAGAVEFNDVKTLLREWITTISDPMEEDILQVVKYCTD  
 LIEEKDLEKLDLVIKYMKRLMQQSVEVWNMAFDLNDVQVVLQQTYGSTLK  
 VT (SEQ ID NO:713); MIKDKGRARTALTSSQPAHLCPENPLLHLKAAVKE  
 KKRNNKKKTIGSPKRIQ (SEQ ID NO:714); KRIQSPLNNKLLNSPAKT  
 LPGACGSPQKLIDGFLKHEGPPAEKPLEELSASTSGVPGLSSLQSDPAGCVRPP  
 15 APNLAGAVEFNDVKTLLREWITTISDPM (SEQ ID NO:715);  
 TISDPMEEDILQVVKYCTDLIEEKDLEKLDLVIKYMKRLMQQSVE  
 SVWNMAFDLNDVQVVLQQTYGSTLKV (SEQ ID NO:716). An additional  
 embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in 12 week embryo and to a lesser extent in  
 20 hemangiopericytoma and frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions, which include, but are  
 not limited to, growth disorders and hemangiopericytoma. Similarly, polypeptides and  
 25 antibodies directed to these polypeptides are useful in providing immunological probes  
 for differential identification of the tissue(s) or cell type(s). For a number of disorders  
 of the above tissues or cells, particularly of the circular and neural system, expression  
 of this gene at significantly higher or lower levels may be routinely detected in certain  
 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,  
 30 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an  
 individual having such a disorder, relative to the standard gene expression level, i.e.,  
 the expression level in healthy tissue or bodily fluid from an individual not having the  
 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID  
 NO: 383 as residues: Leu-4 to Lys-11.

35 The tissue distribution indicates that polynucleotides and polypeptides  
 corresponding to this gene are useful for the treatment of growth disorders,  
 hemangiopericytoma and other soft tissue tumors.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 151**

The translation product of this gene has been found to have homology to a human DNA mismatch repair protein PMS3. Preferred polypeptide fragments comprise the following amino acid sequence: FCHDCKFPEASPAMNCEP (SEQ ID NO:717).  
Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. R95250).

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lymphoma, immunodeficiency diseases, and cancers resulting from genetic instability. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 384 as residues: Met-1 to Lys-6.

The tissue distribution in neutrophils and the sequence homology indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of Hodgkin's lymphoma, since the elevated expression and secretion by the tumor mass may be indicative of tumors of this type. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions.

Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Furthermore, its homology to a known DNA repair protein would suggest gene may be useful in establishing cancer predisposition and prevention in gene therapy applications.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 152**

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, infectious diseases and lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of inflammation and infectious diseases.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 153

One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MASSVPAGGHTRAGGIFLIGKLDLEASLFKSFQWLPPFVLRKKC  
 NFFCWDSSAHSPLHPLSASCSAPACHASDTHLLYPSTRALCPSIFAWLVAPHS  
 20 VFRTNAPGPTPSSQSSPVFPVFPVSFMALIVCXLVCC (SEQ ID NO:720);  
 MASSVPAGGHTRAGGIFLIGKLDLEASLFKSFQWLPPFVLRKKCNFFCWDSSAH  
 SLPLHPLSASCSAPACHA (SEQ ID NO:721);FAWL VAPHSVFRTNAPGPTPS  
 SQSSPVFPVFPVSFMALIVCXLVCC (SEQ ID NO:722). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

25 This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and infectious disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred



epitopes include those comprising a sequence shown in SEQ ID NO: 386 as residues: Ser-11 to Pro-17.

5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of infectious diseases and inflammation.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 154

This gene is expressed in multiple tissues including ovary, uterus, adipose tissue, brain, and the liver.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, uterine, ovarian, brain, and liver cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes  
15 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an  
20 individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnostic or therapeutic uses in  
25 the treatment of the female reproductive system, obesity, and liver disorders, particularly cancer in the above tissues.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 155

This gene maps to chromosome 3, and therefore, may be used as a marker in  
30 linkage analysis for chromosome 3 (See Accession No. D87452).

This gene is expressed in multiple tissues including brain, aortic endothelial cells, smooth muscle, pituitary, testis, melanocytes, spleen, neutrophils, and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological disorders including immunodeficiencies, cancers of the brain and the female reproductive system, as well as cardiovascular disorders, such as

atherosclerosis and stroke. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and immune systems, expression of this gene at  
5 significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution suggest that polynucleotides and polypeptides corresponding to this gene are useful in treatment/detection of disorders in the nervous system, including schizophrenia, neurodegeneration, neoplasia, brain cancer as well as cardiovascular and female reproductive disorders including cancer within the above tissues.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 156

The translation product of this gene shares sequence homology with the human gene encoding cytochrome b561 (See Accession No. P10897). Cytochrome b561 is a transmembrane electron transport protein that is specific to a subset of secretory vesicles  
20 containing catecholamines and amidated peptides. This protein is thought to supply reducing equivalents to the intravesicular enzymes dopamine-beta-hydroxylase and alpha-peptide amidase. Preferred polypeptides of the invention comprise the amino acid sequence:

MAMEGYWRFLALLGSALLVGFLSVIFALVWVLHYREGLGWDGSALEFNWHP  
25 VLMVTGTFVFIQGLAIVYRLPWTWKCSKLLMKSIHAGLNAVAAILAISVVAVFE  
NHNVNNIANMYSLSHSWVGLIAVICYLLQLLSGFSVFLLPWAPLSLRAFLMPIHV  
YSGIVIFGTVIATALMGLTEKLIFSLRDPAYSTFPPEGVFVNTLGLLLLVFGALIF  
WIVTRPQWKRPKEPNSTILHPNGGTEQGARGSMPAYSGNNMDKSDSEL  
NSEVAARKRNALDEAGQRSTM (SEQ ID NO:724); as well as antigenic fragments  
30 of at least 20 amino acids of this gene and/or biologically active fragments. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in anergic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system and metabolism related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product or RNA of this gene is useful for treatment or diagnosis of immune system and metabolic diseases or conditions including Tay-Sachs disease, phenylketonuria, galactosemia, various porphyrias, and Hurler's syndrome.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 157

The translation product of this gene shares sequence homology with collagen which is important in mammalian development. This gene also shows sequence homology with bcl-2. (See Accession No. P80988.) Preferred polypeptide fragments comprise the amino acid sequence: PGRAGPSPGLSLQLPAEPGHPAGNLAPLTSRPQPLCRIPAVPG (SEQ ID NO:725). Also preferred are polynucleotide sequences encoding this polypeptide fragment.

This gene is expressed primarily in HL-60 tissue culture cells and to a lesser extent in liver, breast, and uterus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological diseases, hereditary disorders involving the MHC class of immune molecules, as well as developmental disorders and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and reproductive system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those

comprising a sequence shown in SEQ ID NO: 390 as residues: Ser-39 to Gly-46, Leu-49 to Ala-62.

The tissue distribution and homology to collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hereditary MHC disorders and particularly autoimmune disorders including rheumatoid arthritis, lupus, scleroderma, and dermatomyositis, as well as many reproductive disorders, including cancer of the uterus, and breast tissues.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 158

10 This gene is expressed primarily in the amygdala region of the brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, a variety of brain disorders, particularly those effecting mood and personality. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and/or diagnosis of a variety of brain disorders, particularly bipolar disorder, unipolar depression, and dementia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 159

30 This gene is expressed in a variety of tissues and cell types including brain, smooth muscle, kidney, salivary gland and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers of a variety of organs including brain, smooth muscle, kidney, salivary gland and T-cells and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the central nervous, urinary, salivary, digestive, and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or  
5 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain, smooth muscle, and T-cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of  
10 various neurological, and cardiovascular disorders, but not limited to cancer within the above tissues. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma,  
15 immune deficiency diseases such as AIDS, and leukemia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 160

The translation product of this gene shares sequence homology with collagen which is thought to be important in cellular interactions, extracellular matrix formation,  
20 and has been found to be an identifying determinant in autoimmune disorders. Moreover, this gene shows sequence homology with the yeast protein, Sls1p, an endoplasmic reticulum component, involved in the protein translocation process in Yeast *Yarrowia lipolytica*. (See Accession No. 1052828; see also J. Biol. Chem. 271, 11668-11675 (1996).) With mouse, this same region shows sequence homology with  
25 the heavy chain of kinesin. (See Accession No. 2062607.) Recently, suppression of the heavy chain of kinesin was shown to inhibit insulin secretion from primary cultures of mouse beta-cells. (See Endocrinology 138 (5), 1979-1987 (1997).) Moreover, kinesin was found associated with drug resistance and cell immortalization. (See 468355.) Thus, it is likely that this gene also act as a genetic suppressor elements.

30 This gene is expressed primarily in the greater omentum and to a lesser extent in a variety of organs and cell types including gall bladder, stromal bone marrow cells, lymph node, liver, testes, pituitary, and thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the endocrine, gastrointestinal, and immunological systems, including autoimmune disorders and cancers in a variety of organs and cell types.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 393 as residues: Asn-27 to Leu-47, Gln-81 to Lys-88, Asp-93 to Lys-102, Asn-107 to Leu-116, Met-129 to Glu-141, Glu-150 to Asp-157, Lys-176 to Glu-185, Glu-333 to Tyr-349, Cys-393 to Leu-403, Gln-423 to Gly-429.

The tissue distribution in within various endocrine and immunological tissues combined with the sequence homology to a conserved collagen motif indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of various autoimmune disorders including, but not limited to, rheumatoid arthritis, lupus erythematosus, scleroderma, dermatomyositis. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 161

This gene has homology to the tissue inhibitor of metalloproteinase 2. Such inhibitors are vital to proper regulation of metalloproteins such as collagenases (See Accession No. P16368). In addition, this gene maps to chromosome 17, and therefore, may be used as a marker in linkage analysis for chromosome 17 (See Accession No. P16368).

This gene is expressed primarily in several types of cancer including osteoclastoma, chondrosarcoma, and rhabdomyosarcoma and to a lesser extent in several non-malignant tissues including synovium, amygdala, testes, placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, various types of cancer, particularly cancers of bone and cartilage, as well as various autoimmune disorders. Similarly, polypeptides and antibodies directed

to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the musculoskeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in various cancers and the sequence homology to a collagenase inhibitor indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 162

This gene is homologous to the mitochondrial ATP6 gene and therefore is likely a homolog of this gene family (See Accession No. X76197).

This gene is expressed primarily in brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, a variety of brain disorders, including Down's syndrome, depression, Schizophrenia, and epilepsy. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain tissue indicates this gene is useful for diagnosis of various neurological disorders including, but not limited to, brain cancer.

Additionally the gene product may be used as a target in the immunotherapy of cancer in the brain as well as for the diagnosis of metabolic disorders such as obesity Tay-Sachs disease, phenylketonuria and Hurler's Syndrome.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 163**

This gene is expressed primarily in placenta, neutrophils, and microvascular endothelial cells and to a lesser extent in multiple tissues including brain, prostate, spleen, thymus, and bone.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neutropenia and other diseases of the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in placenta indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis various female reproductive disorders. Additionally the gene product may be used as a target in the immunotherapy of various cancers. Because the gene is expressed in some cells of lymphoid and endocrine origin, the natural gene product may be involved in immune functions and metabolism regulation, respectively. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 164**

This gene is expressed primarily in neutrophils, monocytes, bone marrow, and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system disorders including, but not limited to, autoimmune disorders such as lupus, and immunodeficiency disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders



of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having, such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in various immune system tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of various immunological disorders such as Hodgkin's lymphoma, arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 165

The translation product of this gene shares sequence homology with dystrophin which is thought to be defective in both Duchene and Becker Muscular Dystrophy.

Preferred polypeptide fragments comprise the following amino acid sequence:

MKLLGECSSSIDSVKRLEHKLKEEEESLPGFVNLHSTETQTAGVIDRWELLQAAQ  
 ALSKELRMKQNLQKWQQFNSDLNSIWAUWLGDTTEEELEQLQRLELSTDIQTIELQ  
 IKKLKELQKAVDHRKAILLSINLCSPEFTQADSKESRDLQDRLXQMNGRWDRV  
 CSLLEEWRGLLQDALMQCQGFHEMSHGLLLMLENIDRRKNEIVPIDSNLDAEIL  
 QDHHKQLMQIKHELLESQRLVASLQDMSCQLLVNAEGTDCLEAKEKVHVIGNR  
 LKLLLKEVSRHIKELEKLLDVSSSQQDLSSWSSADELDTSGSVSPXSGRSTPNR  
 QKTPRGKCSLSQPGPSVSSPHSRSTKGGSDSSLSEXPGRSGRGFLFRVLRAA  
 LPLQLLLLLLIGLACL VPMSEEDYSCALSNNFARSFHPMLRYTNGPPPL (SEQ ID  
 NO:726); MKLLGECSSSIDSVKRLEHKLKEEEESLPGFVNLHSTETQTAGVIDR  
 WELLQAAQALSKELRMKQNLQKWQQFNSDLNSIWAUWLGDTTEEELEQLQRLELS  
 TDIQTIELQIK (SEQ ID NO:727); KLKELQKAVDHRKAILLSINLCSPEFTQADSK  
 ESRDLQDRLXQMNGRWDRVCSLLEEWRGLLQDALMQCQGFHEMSHGLLLML  
 ENIDRRKNEIVPIDSNLDAEILQDHHKQLMQIKHELLESQRLVASLQDMSCQL  
 (SEQ ID NO:728); QDMSCQLLVNAEGTDCLEAKEKVHVIGNRLKLLLKEVS  
 RHIKELEKLLDVSSSQQDLSSWSSADELDTSGSVSPXSGRSTPNRQKTPRGKCS  
 LSQPGPSVSSPHS (SEQ ID NO:729); DSSLSEXPGRSGRGFLFRVLRAAL  
 PLQLLLLLLIGLACL VPMSEEDYSCALSNNFARSFHPMLRYTNGPPPL (SEQ ID  
 NO:730). Also preferred are polynucleotide fragments encoding these polypeptide  
 fragments. Furthermore, this gene maps to chromosome 6, and therefore, may be used  
 as a marker in linkage analysis for chromosome 6 (See Accession No. N62896).

This gene is expressed in numerous tissues including the heart, kidney, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, musculoskeletal disorders including Muscular Dystrophy and cardiovascular diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscle tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to dystrophin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of Muscular Dystrophy and other muscle disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 166

This gene is expressed primarily in human cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the central nervous system, including Alzheimer's Disease, Parkinson's Disease, ALS, and mental illnesses. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 399 as residues: Pro-20 to Gly-26, Leu-37 to Pro-42, His-57 to Gly-63.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of diseases of the central nervous system and may protect or

enhance survival of neuronal cells by slowing progression of neurodegenerative diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 167

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MKLLICGNYLAPSHSESSRRCCLLCFYPLCLEINFGMKVFLSMPFLVLFQ

SLIQED (SEQ ID NO:731). Polynucleotides encoding such polypeptides are also provided. This gene is believed to reside on chromosome 15. Therefore polynucleotides  
10 derived from this gene are useful in linkage analysis as chromosome 15 markers.

This gene is expressed primarily in human testes tumor and to a lesser extent in normal human testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
15 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the testes, particularly cancer, and other reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of  
20 the male reproductive tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily  
25 fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of testicular diseases including cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 168

30 This gene is expressed primarily in fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, conditions affecting hematopoietic development and metabolic diseases.  
35 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the

hepatic system, and fetal hematopoietic system. expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level. i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 401 as residues: His-7 to Trp-17, Leu-19 to Lys-27, Pro-33 to Gly-44, Lys-68 to Gly-74, Lys-85 to Cys-95.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of diseases of the developing liver and hematopoietic system, and act as a growth differentiation factor for hematopoietic stem cells.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 169

The polypeptide encoded by this gene is believed to be a membrane bound receptor. The extracellular domain of which is expected to consist of the following amino acid sequence:

RILLVKYSANEENKYDYLPTTFVNVCSSELVKLVFCVLVSFCVIKKDHQSRNLKY  
 ASWKEFSDFMKWSIPAFLYFLDNLIVFYVLSYLQPAMAVIFS NFSIITTALLFRIV  
 LKXRLNWIQWASLLTLFLSIVALTAGTKTLQHNLAGRGFHHDAFFSPNSCLL  
 FRNECPRKDNCTAKEWTFPEAKWNTTARVFSHIRLGMGHVLIIVQCFISSMANI  
 YNEKILKEGNQLTEXIFIQNSKLYFFGILFNGLTLGLQRSNRDQIKNCGFFYGH  
 S (SEQ ID NO:732). Thus, preferred polypeptides encoded by this gene comprise the extracellular domain as shown above. It will be recognized, however, that deletions of either end of the extracellular domain up to the first cysteine from the N-terminus and the first cysteine of the C-terminus, is expected to retain the biological functions of the full-length extracellular domain because the cysteines are thought to be responsible for providing secondary structure to the molecule. Thus, deletions of one or more amino acids from either end (or both ends) of the extracellular domain are contemplated. Of course, further deletions including the cysteines are also contemplated as useful as such polypeptides is expected to have immunological properties such as the ability to evoke and immune response. Polynucleotides encoding all of the foregoing polypeptides are provided.

This gene is expressed primarily in human osteoclastoma and to a lesser extent in hippocampus and chondrosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, cancers, particularly those of the bone and connective tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 402 as residues: Met-1 to Cys-6, Ala-41 to Tyr-49, Lys-76 to Lys-84.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis of cancers of the bone and connective tissues, and may act as growth factors for cells involved in bone or connective tissue growth.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 170

Preferred polypeptides encoded by this gene comprising the following amino acid sequence:

NSVPNLQTLAYL TEAIGPEPAIPRXPREPPVATSTPATPSAGPQPLPTGTV  
LVPGGPAPPCLGEAWALLPPCRPSLTSCFWSPRPSPWKETGV (SEQ ID NO:733). Polynucleotides encoding such polypeptides are also provided herein.

This gene is expressed primarily in hematopoietic progenitor cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the blood including cancer and autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the blood/circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 403 as residues: Gln-4 to His-10, Pro-25 to His-32.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis of diseases involving growth differentiation of hematopoietic cells.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 171

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequences: ALQLAFYPDAVEEWLEENVHPSLQRLQXLLQDLSEVVSAPP (SEQ ID NO:734); and/or CHPPALAGTLLRTPEGRAHARGLLLEAGGA (SEQ ID NO:735). Polynucleotides encoding such polypeptides are also provided. The protein product of this gene shares sequence homology with metallothionines. Thus, polypeptide encoded  
10 by this gene are expected to have metallothionine activity, such activities are known in the art and described elsewhere herein.

This gene is expressed primarily in kidney cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
15 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the kidney including cancer and renal dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the renal system, expression of this gene at significantly higher or lower levels may be routinely detected  
20 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID  
25 NO: 404 as residues: Ser-47 to Gln-52.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the kidney including kidney failure.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 172

This gene is expressed primarily in 12 week old early stage human.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing embryo, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 405 as residues: Gln-31 to Thr-43, Gly-51 to Ser-58, Pro-65 to Pro-72.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of developmental problems with fetal tissue. The gene may be involved in vital organ development in the early stage, especially hematopoiesis, cardiovascular system, and neural development.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 173

The translation product of this gene shares sequence homology with TGN38, an integral membrane protein previously shown to be predominantly localized to the trans-Golgi network (TGN) of cells.

20 This gene is expressed primarily in developing embryo and to a lesser extent in cancer tissues including lymphoma, endometrial, prostate and colon.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 406 as residues: His-65 to Ser-72, Pro-82 to Gly-91, Pro-98 to Glu-118, Ser-126 to Gly-166, Pro-180 to Asp-188, Tyr-209 to Lys-214, Gln-220 to Leu-228.

35 The tissue distribution and homology to an integral membrane protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for

diagnosis of cancers and developmental abnormalities where aberrant expression relates to an abnormality.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 174

5       The translation product of this gene shares sequence homology with a dnaJ heat shock protein from *E. coli* which is allelic to sec63, a gene that affects transit of nascent secretory proteins across the endoplasmic reticulum in yeast.

      This gene is expressed primarily in Hodgkin's lymphoma and to a lesser extent in testes.

10       Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification  
15       of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to  
20       the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 407 as residues: Thr-13 to Trp-21, Arg-74 to Asp-81.

      The tissue distribution and homology to dnaJ indicates that polynucleotides and  
25       polypeptides corresponding to this gene are useful as a diagnostic for cancer including Hodgkin's lymphoma.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 175

30       This gene is expressed primarily in endothelial cells and to a lesser extent in bone marrow stromal cells.

      Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases involving angiogenic abnormalities including diabetic  
35       retinopathy, macular degeneration, and other diseases including arteriosclerosis and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell



type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treating diseases where an increase or decrease in angiogenesis is indicated and as a factor in the wound healing process.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 176

The translation product of this gene shares sequence homology with MAT8 (mouse) which is thought to be important in regulating chloride conductance in cells (particularly in the breast) by modulating the response mediated by cAMP and protein kinase C to extracellular signals.

This gene is expressed primarily in amniotic cells and hematopoietic cells including macrophages, Neutrophils, T cells, TNF induced aortic endothelium and to a lesser extent in testes, TNF induced epithelial cells, and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory responses mediated by T cells, macrophages, and/or neutrophils particularly those involving TNF, and also cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 409 as residues: Thr-19 to Ala-33, Leu-54 to Asp-82, Pro-89 to Ala-97, Pro-100 to Lys-125, Ser-127 to Phe-135, Gly-164 to Leu-169, Cys-173 to Arg-178.

The tissue distribution and homology to mat-8 indicates that polynucleotides and polypeptides corresponding to this gene are useful for modifying inflammatory

responses to cytokines such as TNF and thus modifying the duration and/or severity of inflammation. Polynucleotides and polypeptides derived from this gene are thought to be useful in the diagnosis and treatment of cancer.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 177

This gene is expressed primarily in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, vascular restenosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases associated with vascular response to injury such as vascular restenosis following angioplasty..

## FEATURES OF PROTEIN ENCODED BY GENE NO: 178

One embodiment of the claimed invention comprises:

MRPDWKAGAGPGGPPQKPAPSSQRKPPARPSAAAAIAVAAAEERRLRQRN  
RLRLEEDKPAVERCLEELVFGDVENEDALLRRLRGPRVQEHEDSGDSEVENEAK  
KGNFPPQKKPVWVDEEDEDEEMVDMMNRRFRKDMMKNAESKLSKDNLKK  
RLKEEFQHAMGGVPAWAETTKRKTSSDDESEEDDDLLQRTGNFISTSTSLPRG  
ILKMKNQCQHANAEPTVARISICAVPSRCTDCDGCWD (SEQ ID NO:737); or  
CLEELVFGDVENEDALLRRLRGPRVQEHEDSGDSEVENEAKGNFPPQKKPV  
WVDEEDEDEEMVDMMNRRFRKDMMKNAESKLSKDNLKKRLKEEFQHAMG  
GVPWAETTKRKTSSDDESEEDDDLLQRTGNFISTSTSLPRGILKMKNQCQHA  
NAERPTVARISICAVPSRCTDCDGC (SEQ ID NO: 738). LKEKIVRSFEVSPDGS  
FLLINGLAGYLHLLAMKTKELIGSMKINGRVAASTFSSDSKKVYASSGDGEVYV  
WDVNSRKCLNRFVDEGSYGLSLATSRNGQYVACGSNCGVVNIYNQDSCLQE  
TNPKPIKAIMNLVTGVTSLSLTFNPTTEILAIASEKMKEAVRLVHLPSTVFSNFPVI  
KNKNISHVHTMDFSPRSGYFALGNEKGKALMYRLHHYSDF (SEQ ID NO:739);

and/or KINGRVAASTFSSDSKKVYASSGDGEVYVWDVNSRKCLNRFVDEGSL  
YGLSLATSRNGQYVACGSNCGVVNIYNQDSCLQETNPKPIKAIMNLVTGVTSLT  
FNPTTEILAIASEKMKEAVRLVHLPSCCTVFSNFPVIKNKNISHVHTMDFSPRSG  
YFALGNEKGKAL (SEQ ID NO:740).

- 5 This gene is expressed primarily in epididymus and endometrial tumors and to a lesser extent in T cell lymphoma and cell lines derived from colon cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of the reproductive organs including testis and endometrial cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 411 as residues: Ser-67 to Lys-72, Val-87 to Leu-93, Tyr-128 to Pro-141, Asp-204 to Gly-210.

The tissue distribution indicates that the protein products of this gene are useful for treating tumors of the endometrium or epithelial tumors of the reproductive system.

## 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 179

Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MRILQLILLALATGLVGGETRIIKGFECKLHSQPWQAALFEKTRLLCGATLIAPR  
WLLTAAHCLKPRYIVHLGQHNLQKEEGCEQTRTATESFPHPGFNNSLPNKDH  
30 RNDIMLVKMASPVSITWAVRPLTLSSRCVTAGTSCSFPAGAARPDPSYACLTPC  
DAPTSPSLSTRSVRTPTPATSQTPWCVPACRKGARTPARVTPGALWSVTSLFKA  
LSPGARIRVRSPESLVSTRKSANMWTGSRRR (SEQ ID NO:741); ETRIIKGFEC  
KLHSQPWQAALFEKTRLLCGATLIAPRWLLTAAHCLKPRYIVHLGQHNLQKEE  
GCEQTRTATESFPHPGFNNSLPNKDHRNDIMLVKMASPVSITWAVRPLTLSSR  
35 CVTAGTSCSFPAGAARPDPSYACLTPCDAPTSPSLSTRSVRTPTPATSQTPWCVP  
ACRKGARTPARVTPGALWSVTSLFKALSPGARIRVRSPESLVSTRKSANMWTG

SRRR (SEQ ID NO:742); or CKLHSQPWQAALFEKTRLLCGATLIAPRWLLT  
AAHCLKPRYIVHLGQHNQKEEGCEQTRTATESFPHPGFNS

(SEQ ID NO:743). The translation product of this gene shares sequence homology  
with neuropsin a novel serine protease which is thought to be important in modulating  
extracellular signaling pathways in the brain. Owing to the structural similarity to other  
serine proteases the protein products of this gene are expected to have serine protease  
activity which may be assayed by methods known in the art and described elsewhere  
herein.

This gene is expressed primarily in endometrial tumor and to a lesser extent in  
colon cancer, benign hypertrophic prostate; and thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
not limited to, cancers of the endometrium or colon and benign hypertrophy of the  
prostate. Similarly, polypeptides and antibodies directed to these polypeptides are  
useful in providing immunological probes for differential identification of the tissue(s)  
or cell type(s). For a number of disorders of the above tissues or cells, particularly of  
the urogenital or reproductive systems, expression of this gene at significantly higher or  
lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded  
tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or  
another tissue or cell sample taken from an individual having such a disorder, relative to  
the standard gene expression level, i.e., the expression level in healthy tissue or bodily  
fluid from an individual not having the disorder. Preferred epitopes include those  
comprising a sequence shown in SEQ ID NO: 412 as residues: Gly-12 to Ser-22, Pro-  
34 to Ser-53.

The tissue distribution and homology to serine proteases indicates that  
polynucleotides and polypeptides corresponding to this gene are useful for diagnosing  
or treating hyperproliferative disorders such as cancer of the endometrium or colon and  
hyperplasia of the prostate.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 180

Preferred polypeptide encoded by this gene comprise the following amino acid  
sequence: VLQGRYFSPILEMRRLRPEGXXNLPGGSRAQKEPRQDLTLVLWPHC  
PHFAMTRSYPVKQCMVQGSFYCIFKGPVQNWNC (SEQ ID NO:744).

Polynucleotides encoding such polypeptide are also provided.

This gene is expressed primarily in fetal brain

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, identifying and expanding stem cells in the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for detecting and expanding stem cell populations in the (or of the) central nervous system.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 181**

This gene is expressed primarily in early stage human brain and a stromal cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities of the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 414 as residues: Gln-42 to Gln-47, Gln-54 to Pro-60.

The tissue distribution indicates that the protein products of this gene play a role in the development of the central nervous system. Therefore this gene and its products

are useful for diagnosing or treating developmental abnormalities of the central nervous system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 182

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MPIIDQVNPELHDFMQSAEVTIFALSWLITWFGHVLSDFRHVVRLYDF  
FLACHPLMPIYFAAVIVLYREQEVLDCDCDMASVHHLLSQIPQDLPYETLISRXE  
TFLFSFHPNLLGRPLNSKLRGRQPLLSKTLSTWHQPSRGLIWCCGSGXRGLL  
10 RPEDRTKDVLTTPRTNRFVKLAVMGLTVALGAAALAVVKSALWAPKFQLQL  
FP (SEQ ID NO:745); or CPEFFIPATLPCPFVFAFTSEASSRAYLTQRGPGGLAQ  
NLMPLPVGFWMGSLPPPWCWRKWVSEACSCFC (SEQ ID NO:746) These  
polypeptides are structurally similar to various TGF-beta family members. Thus, this  
polypeptide is expected to have a variety of activities in the modulation of cell growth  
15 and proliferation.

This gene is expressed primarily in osteoclastoma, microvascular endothelium, and bone marrow derived cell lines.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematological diseases particularly involving aberrant proliferation of stem cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of  
25 the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
30 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 415 as residues: Ser-33 to Ala-39.

The tissue distribution indicates that the protein products of this gene is useful for treating disorders of the progenitors of the immune system. Applications include in vivo expansion of progenitor cells, ex vivo expansion of progenitor cells, or the  
35 treatment of tumors of the circulatory system, such as lymphomas.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 183

This gene maps to chromosome 17 and therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 17. In specific embodiments, polypeptides of the invention comprise the sequence:

- 5 GFGSVSAAGRSGGTWQPVQ (SEQ ID NO:747); PGGLAVGSRWWSRSLT (SEQ ID NO:748); LEPSRQRRPRRRGGTSRPETDQRAKCWRQL (SEQ ID NO:749); and/or VCLRCQNRMEN (SEQ ID NO:750). In further specific embodiments, polypeptides of the invention comprise the sequence: MAACTARRPGR GQPLVVPVADXGPVAKAALCAAXAGAFSPASTTTTTRHLSSRNRPPEGKVLETV
- 10 GVFEVPKQNGKYETGQLFLHSIFGYRGVVLPWQARLXDRDVASAAPEKAEN PAGHGSKEVKGKTHTYYQVLIDARDCPHISQRSQTEAVTFLANHDDSRALYAIP GLDYVSHEDILPYTSTDQVPIQHELPERFLLYDQTKAPPFVARETLRAWQEKNH PWLELSDVHRETTENIRVTVIPFYMGMREAQNSHVYWWRYCIRLENLSDSDVVQ LRERHWRIFSLSGTLETVRGRGVVGREPVLSKEQPAFQYSSHVSLQASSGHMW
- 15 GTFRFERPDGSHFDVRIPPFSLESNKDEKTPPSGLHW (SEQ ID NO:751); MAACTARRPGRGQPLVVPVADXGPVAKAALCAA (SEQ ID NO:752); VLETVGVFEVPKQNGKYETGQLFLHSIFGYRGVV (SEQ ID NO:757); GLDYVSHEDILPYTST (SEQ ID NO:758); DVHRETTENIRVTVIPFYM (SEQ ID NO:759); WWRYCIRLENLSDSDVVQLRER (SEQ ID NO:760); and/or PAFQYSS
- 20 HVSLQASSGHMWGTFRFER (SEQ ID NO:761). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in gall bladder, prostate, and fetal brain, and to a lesser extent in a few tumor and fetal tissues.

- Therefore, polynucleotides and polypeptides of the invention are useful as
- 25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, growth related disorders such as cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders
- 30 of the above tissues or cells, particularly of the prostate, gall bladder, and fetal brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,
- 35 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of growth-related disorders, such cancers.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 184

In specific embodiments, polypeptides of the invention comprise the sequence: SLCCPEGAEGC (SEQ ID NO:762) and/or QLKKTHYDRPCP (SEQ ID NO:763). Polynucleotides encoding these polypeptides are also encompassed by the invention.

10 This gene is expressed primarily in stromal cell, tonsil, and glioblastoma and to a lesser extent in some other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and inflammatory disorders and glioblastoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the stromal cells, tonsil, and glioblastoma expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Additionally, it is believed that the product of this gene regulates pancreatic cell differentiation into beta cells. Accordingly, polynucleotides and polypeptides of the invention are useful in the treatment of insulin-dependent diabetes mellitus and associated conditions e.g. pancreatic hypofunction and the prevention, as well as the treatment of undifferentiated type pancreatic cancers. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 417 as residues: Pro-27 to Ala-32.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune and inflammatory disorders and glioblastoma.

## 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 185

This gene is expressed primarily in hepatocellular carcinoma and to a lesser extent in other tissues.



Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, liver diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 418 as residues: Gly-32 to Lys-39.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 186

This gene is expressed primarily in hippocampus and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hippocampus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 187

This gene is expressed primarily in bone cancer and hippocampus and to a lesser extent in osteoclastoma and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bone-related disorders and neuronal diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, osteoclast, and hippocampus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of bone-related disorders and neuronal diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 188

This gene maps to chromosome 4 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 4.

This gene is expressed primarily in neuronal tissues such as hippocampus, spinal cord, and hypothalamus and to a lesser extent in a few other tissues such as ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 189**

This gene maps to chromosome 10, therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 10.

5 This gene is expressed primarily in neuronal tissues and immune tissues, and to a lesser extent in a few other tissues such as skin tumor, lung etc.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal and immune-related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal and immune-related tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 422 as residues: Pro-19 to Asp-25.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal and immune-related disorders.

**25 FEATURES OF PROTEIN ENCODED BY GENE NO: 190**

The translation product of this gene shares sequence homology with human N33, a gene located in a homozygously deleted region of human metastatic prostate cancer which is thought to be important in prevention of prostate cancer. In specific embodiments, polypeptides of the invention comprise the sequence:

30 AQRKKEMVLSEKVSQ LMEWTNKR PVIRMGDKFRRL VKAPPRNYSVIVMFTA  
LQLHRQCVVCKQADEEFQILANSWRYSSAFTNRIFFAMVDFDEGSDVFQMLNM  
NSAPTFINFPAKGKPKRGDTYELQVRGFSAEQLARWIADRTDVNIRVIRPPNMA  
ARWRFWCVSVT (SEQ ID NO:765); MVVALLIVCDVPSAS (SEQ ID NO:766);  
AQRKKEMVLSEKVSQ L (SEQ ID NO:767); MEWTNKR PVIRMGDKF (SEQ  
35 ID:768); RRLVKAPPRNYSVIVMFTALQLHRQCVVCKQADEEFQILANSWRY  
SSAFTNRIFFA (SEQ ID NO:769); MVD FDEGSDVFQMLNMNSAPTFINFPAK  
GKP (SEQ ID NO:770); KRGDTYELQVRGFSAEQLARWIADRTDVNIRVIRPPN-

(SEQ ID NO:771); and/or YAGPLMLGLLLA VIGGLVYLRRVTWNFSLIKLDGLLQL CVLCLL (SEQ ID NO:772). Polynucleotides encoding these polypeptides are also encompassed by the invention.

5 This gene is expressed primarily in infant adrenal gland prostate cell line and to a lesser extent in a few other tissues like liver, smooth muscle etc.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate cancer and endocrine disorders. Similarly, polypeptides and  
10 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate and adrenal gland, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,  
15 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 423 as residues: Pro-34 to Gly-43, Arg-113 to Pro-120.

20 The tissue distribution and homology to N33 indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for prostate cancer and endocrine disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 191

25 This gene is expressed primarily in T cell and to a lesser extent in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to  
30 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal  
35 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue

or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 424 as residues: Trp-3 to Phe-9.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders.

5

## FEATURES OF PROTEIN ENCODED BY GENE NO: 192

This gene maps to chromosome 6, therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 6. Neural activity and neurotrophins induce synaptic remodeling in part by altering gene expression. This  
10 gene is believed to be a glycosylphosphatidylinositol-anchored protein encoded by a hippocampal gene and to possess neural activity. This molecule is believed to be expressed in postmitotic-differentiating neurons of the developing nervous system and neuronal structures associated with plasticity in the adult. Message of this gene is believed to be induced by neuronal activity and by the activity-regulated neurotrophins  
15 BDNF and NT-3. The product of this gene is believed to stimulate neurite outgrowth and arborization in primary embryonic hippocampal and cortical cultures and to act as a downstream effector of activity-induced neurite outgrowth. In specific embodiments, polypeptides of the invention comprise the sequence: DAVFKGFSDCLLKLGD (SEQ ID NO:773); CQEGAKDMWDKLRKESKNLN (SEQ ID NO:774);  
20 VLLVLSAALATWLSF (SEQ ID NO:775); MGLKLNGRYISLILAVQIAYLVQAVR AAGKCD VFKGFSDCLLKLGD (SEQ ID NO:776); PAAWDDKTNIKTVCTYW EDFHSCTVTALTDCQEGAKDMWDKLRKESKNLNIQGSLFELCGSGNGAAGSL LPAPVLLVLSAALATWLSF (SEQ ID NO:777); and/or MGLKLNGRYISLILAVQIAYLVQAVRAAGKCD VFKGFSDCLLKLGDXXXXXXXXPAAWDDKTNIKTVCTY  
25 TYWEDFHSCTVTALTDCQEGAKDMWDKLRKESKNLNIQGSLFELCGSGNGAAGSLLPAPVLLVLSAALATWLSF (SEQ ID NO:778). Polynucleotides encoding this polypeptide are also encompassed by the invention.

This gene is expressed primarily in human placenta, endometrial tumor and tissues of the central nervous system (CNS).

30

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, relating to reproductive disorders, cancers and neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
35 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and neurological disorders, expression of this gene at significantly higher

or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 425 as residues: Asp-47 to Asp-63, His-75 to Tyr-80, Pro-83 to Tyr-89.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of reproductive disorders such as endometrial tumors. Expression of this gene in tissues of the CNS and its strong homology to Neuritin suggest that the protein product from this gene may also be used in the treatment and diagnosis of neurological disorders and in the regeneration of neural tissues, e.g., following injury.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 193**

The translation product of this gene shares sequence homology with tenascin which is thought to be important in development. The translation product of this gene is believed to be a ligand of the fibroblast growth factor family. FGF ligand activity is known in the art and can be assayed by methods known in the art and disclosed elsewhere herein.

This gene is expressed primarily in endometrial tumors, and other types of tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 426 as residues: Gly-29 to Glu-34, Arg-71 to Arg-76, Thr-176 to Cys-182, Gly-184 to Glu-199.

The tissue distribution and homology to tenascin indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 194

In specific embodiments, polypeptides of the invention comprise the sequence: MNSAAGFSHLDRRERVVLKLGESFEKQPRCASTLC (SEQ ID NO:779).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in fetal human lung and neutrophils.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lung development and respiratory disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes  
15 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the respiratory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual  
20 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in fetal lung and neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis  
25 and treatment of lung and immunity related diseases, for example, lung cancer, viral, fungal or bacterial infections (e.g. lesions caused by tuberculosis), inflammation (e.g. pneumonia), metabolic lesions etc.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 195

30 This gene is expressed primarily in breast lymph node.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunal disorders. Similarly, polypeptides and antibodies directed to  
35 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immunal disorders.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 196

This gene maps to chromosome 5 and accordingly, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 5. The translation product of this gene shares sequence homology with human M-phase phosphoprotein 4 which is thought to be important in phosphorylation and signal transduction processes. In specific embodiments, polypeptides of the invention comprise the sequence:

15 TIYPTEELQAVQKIVSITERALKLVSD (SEQ ID NO:780); RALKGVLRV  
GVLAAGLLLRGDRNVNLVLLC (SEQ ID NO:781); ALAALRHAKWFQARAN  
GLQSCVIRILRDLQCRVPTWS (SEQ ID NO:782); GDALRRVFECISSGIL (SEQ  
ID NO:783); LAFRQIHKVLGMDPLP (SEQ ID NO:784); and/or TIYPTEELQAVQ  
20 KIVSITERALKLVSDSLSEHEKNKNKEGDDKKEGGKDRALKGVLRVGVLAAG  
LLLRGDRNVNLVLLCSEKPSKTLRSIAENLPKQLAVISPEKYDIKCAVSEAAIIL  
NSCVEPKMQVTITLTSPIREENMREGDVTSGMVKDPPDVLDRQKCLDALAALR  
HAKWFQARANGLQSCVIRILRDLQCRVPTWSDFPSWAMELLVEKAISSASSP  
QSPGDALRRVFECISSGILKGSPGLLDPCCKDPFDLATMTDQQREDITSSAQFA  
25 LRLLAFRQIHKVLGMDPLPQMSQRFNIHNNRKRRRDSDGVDGFEGGKDKK  
DYDNF (SEQ ID NO:785). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in Human Hippocampus and to a lesser extent in Prostate, Human Frontal Cortex.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders related to reproductive system and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing  
35 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system and nervous system, expression of this gene at significantly higher or lower



levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to human M-phase phosphoprotein 4 indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of reproductive and nervous system disorders.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 197

In specific embodiments, polypeptides of the invention comprise the sequence: MGSQHSAAARPSSCRRKQEDDRDG (SEQ ID NO:786); LLAEREQEEAIAQFPYVEFTGRDSITCLTC (SEQ ID NO:787); and/or QGTGYIPTEQVNELVALIPHSDQRLRPQRTKQYV (SEQ ID NO:788).

15 Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in Human Primary Breast Cancer and to a lesser extent in Human Adult Spleen, Hodgkin's Lymphoma I, Salivary Gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and immunal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 430 as residues: Ser-126 to Gly-138.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and immunal disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 198**

This gene is expressed primarily in monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, blood cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of blood cell disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 199**

This gene is expressed primarily in Human Ovary and Synovia and to a lesser extent in Human 8 Week Whole Embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and developmental system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of reproductive and developmental disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 200**

This gene maps to chromosome 8 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 8. The translation product of this gene shares limited sequence homology with collagen proline rich domain.

This gene is expressed primarily in CNS.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 433 as residues: Pro-35 to Asp-41.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological diseases.

**25 FEATURES OF PROTEIN ENCODED BY GENE NO: 201**

Translation product of this gene shares homology with a mammalian histone H1a protein. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: ARLNVGRESLKREMLKSQGVKVSSESPMGAR HSSWPEGAAFCCKKVQGAQMCFPPRR (SEQ ID NO:789); ARLNVGRESLKR EML (SEQ ID NO:790); LKSQGVKVSSESPMGARHSSW (SEQ ID NO:791); AFCKKVQGAQMCFPPRR (SEQ ID NO:792). An additional embodiment is the polynucleotide fragments encoding these polypeptide (See Accession No. pirlS24178) fragments.

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at

5 significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in vital immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such

15 as AIDS, and leukemia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 202

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as

20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential

25 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level

30 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for

35 immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 203**

This gene is expressed primarily in Neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as  
5 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
not limited to, infectious disorders, immune disorders, and cancers. Similarly,  
polypeptides and antibodies directed to these polypeptides are useful in providing  
immunological probes for differential identification of the tissue(s) or cell type(s). For  
10 a number of disorders of the above tissues or cells, particularly of the immune system,  
expression of this gene at significantly higher or lower levels may be routinely detected  
in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,  
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from  
an individual having such a disorder, relative to the standard gene expression level, i.e.,  
15 the expression level in healthy tissue or bodily fluid from an individual not having the  
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID  
NO: 436 as residues: Thr-31 to Lys-36.

The tissue distribution indicates that polynucleotides and polypeptides  
corresponding to this gene are useful for diagnosis and treatment of infectious  
20 disorders, immune disorders, and cancers. Since the gene is expressed in cells of  
lymphoid origin, the natural gene product may be involved in immune functions.  
Therefore it may be also used as an agent for immunological disorders including  
arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Protein, as  
well as, antibodies directed against the protein may show utility as a tumor marker  
25 and/or immunotherapy targets for the above listed tumors and tissues.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 204**

This gene maps to chromosome 16 and therefore polynucleotides of the  
invention can be used in linkage analysis as markers for chromosome 16. The  
30 translation product of this gene shares sequence homology with lactate dehydrogenase  
which is thought to be important in lactate metabolism.

This gene is expressed primarily in human tonsils and to a lesser extent in  
Spleen, and Neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as  
35 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions, which include, but are  
not limited to, immune disorders, infectious disorders, and cancers. Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune disorders, infectious disorders, and cancers, expression of this gene at significantly  
5 higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue  
10 or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 437 as residues: Gly-7 to Ser-12.

The tissue distribution and homology to lactate dehydrogenase gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, infectious disorders, and cancers.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 205

The translation product of this gene shares sequence homology with Gcap1 protein which is developmentally regulated in brain.

This gene is expressed primarily in placenta and endometrial tumor and to a lesser extent in several other tumors.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, vasculogenesis/angiogenesis and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological  
25 probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system and tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from  
30 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to Gcap1 protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis  
35 and treatment of disorder or dysfunction of vascular system of tumorigenesis.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 206**

In specific embodiments, polypeptides of the invention comprise the sequence  
MPYAQWLAENDRFEEAQKAFHKAGRQREA (SEQ ID NO:799);

VQVLEQLTNNAVAESRFNDAAYYYWMLSMQCLDIAQD (SEQ ID NO:794);

5 PAQKDTMLGKFYHFQRLAELYHGYHAIHRHTEDP (SEQ ID NO: 795);

FSVHRPETLFNISRFLHSLPKDTPSGISKVKILFT (SEQ ID NO:800);

LAKQSKALGAYRLARHAYDKLRGLYIP (SEQ ID NO:796); ARFQKSIELG

TLTIRAKPFHDSEELVPLCYRCSTNN (SEQ ID NO: 797); and/or PLLNNLGNVC

INCRQPFISSASSYDVLHLVEFYLEEGITDEEAISLIDLEVLRPKRDDRQLEICKQQ

10 LPDSCG (SEQ ID NO:798). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

15 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive and endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells; particularly of the reproductive and endocrine systems,

20 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the  
25 disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of male reproductive and endocrine disorders.

**30 FEATURES OF PROTEIN ENCODED BY GENE NO: 207**

This gene is expressed in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are  
35 not limited to, lung diseases such as cystic fibrosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the respiratory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 440 as residues: Tyr-49 to Cys-54.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of disorders associated with developing lungs particularly in premature infants where the lungs are the last tissues to develop. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of lung tumors since the gene may be involved in the regulation of cell division, particularly since it is expressed in fetal tissue. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.



Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
1	HLHDS67	97979 03/27/97	Uni-ZAP XR	11	2526	427	2526	458	458	234	1	30	31	30
2	HLHDZ58	97979 03/27/97	Uni-ZAP XR	12	1131	1	1131	129	129	235	1	14	15	115
3	HLMMJ13	97979 03/27/97	Lambda ZAP II	13	941	39	941	62	62	236	1	44	45	102
3	HLMMJ13	97979 03/27/97	Lambda ZAP II	218	941	39	941	245	245	441	1	35	36	41
4	HLTEI25	97979 03/27/97	Uni-ZAP XR	14	843	1	843	155	155	237	1	19	20	42
5	HMSJX24	97979 03/27/97	Uni-ZAP XR	15	1018	1	1018	90	90	238	1	18	19	36
6	HNFE65	97979 03/27/97	Uni-ZAP XR	16	661	1	661	76	76	239	1	28	29	127
7	HNHDX07	97979 03/27/97	Uni-ZAP XR	17	553	1	553	106	106	240	1	23	24	66
8	HNHGC82	97979 03/27/97	Uni-ZAP XR	18	869	1	869	101	101	241	1	21	22	68
9	HNHGO09	97979 03/27/97	Uni-ZAP XR	19	959	1	959	176	176	242	1	21	22	44
10	HOUBE18	97979 03/27/97	Uni-ZAP XR	20	1446	1	1446	101	101	243	1	27	28	50
11	HOUDL69	97979 03/27/97	Uni-ZAP XR	21	1471	579	1460	692	692	244	1	31	32	42

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
12	HPMF171	97979 03/27/97	Uni-ZAP XR	22	1402	242	1402	401	401	245	1	32	33	60
13	HPMGQ55	97979 03/27/97	Uni-ZAP XR	23	1047	1	1047	164	164	246	1	26	27	35
14	HPQAC69	97979 03/27/97	Lambda ZAP II	24	990	1	988	82	82	247	1	20	21	37
15	HPTBB03	97979 03/27/97	Uni-ZAP XR	25	1208	350	1173	398	398	248	1	29	30	210
16	HPTWA66	97979 03/27/97	pBluescript	26	1922	1381	1922	24	24	249	1	33	34	547
16	HPTWA66	97979 03/27/97	pBluescript	219	575	1	575	148	148	442	1	22	23	65
17	HPTWC08	97979 03/27/97	pBluescript	27	1951	1422	1874	219	219	250	1	19	20	299
18	HRGCZ46	97979 03/27/97	Uni-ZAP XR	28	3989	2635	3989		2748	251	1	16	17	39
19	HSADVU34	97979 03/27/97	Uni-ZAP XR	29	3735	2966	3735	272	272	252	1	30	31	594
19	HSADVU34	97979 03/27/97	Uni-ZAP XR	220	3018	1929	3018	26	26	443	1	1	2	156
20	HSDFW61	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	30	1667	59	1625	138	138	253	1	32	33	130
21	HSDGP60	97974 04/04/97	Uni-ZAP XR	31	1408	1	1408	285	285	254	1			20

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		209080 05/29/97												
22	HSOAJ55	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	32	2031	1273	2031	1285	1285	255	1	29	30	30
23	HSQEO84	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	33	971	13	971	91	91	256	1	19	20	218
23	HSQEO84	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	221	968	8	968	86	86	444	1	20	21	56
24	HSXAM05	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	34	1792	369	1792	470	470	257	1	26	27	49
25	HSXAS67	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	35	896	1	896	96	96	258	1	32	33	121
26	HTDAF28	97974 04/04/97 209080 05/29/97	pSport1	36	912	1	912	38	38	259	1	22	23	87
27	HTEGQ64	97974	Uni-ZAP XR	37	1382	67	1382	271	271	260	1			25

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		04/04/97 209080 05/29/97												
28	HTGEU09	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	38	872	1	872	74	74	261	1	18	19	28
29	HTOAM21	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	39	812	1	812	41	41	262	1	30	31	43
30	HTPBW79	209511 12/03/97	Uni-ZAP XR	40	1515	118	1507	302	302	263	1	24	25	362
30	HTSEV09	97974 04/04/97 209080 05/29/97	pBluescript	222	1404	1	1265	92	92	445	1	19	20	415
31	HJPCD40	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	41	704	22	704		117	264	1	18	19	127
32	HTWBY48	97974 04/04/97 209080 05/29/97	pSport1	42	1094	1	1094	32	32	265	1	34	35	53
33	HTWC146	97974 04/04/97	pSport1	43	1821	892	1647	56	56	266	1	26	27	28

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		209080 05/29/97												
34	HTXG175	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	44	1024	30	1024		167	267	1	20	21	25
35	HWTF59	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	45	983	779	983	85	85	268	1	30	31	221
35	HWTF59	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	223	707	488	707	514	514	446	1	41	42	64
36	HADA574	97974 04/04/97 209080 05/29/97	pSport1	46	2421	664	1587	710	710	269	1			2
37	HAGFB60	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	47	840	1	840	97	97	270	1	30	31	48
38	HATFE60	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	48	2432	1193	2246	1491	1491	271	1	17	18	51
39	HBM5N25	97974	Uni-ZAP XR	49	1742	1165	1742	1207	1207	272	1	23	24	31

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
40	HCDAR68	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	50	1487	181	1455	325	325	273	1	35	36	56
41	HCE3J79	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	51	1328	251	1328	525	525	274	1			21
42	HMDAN54	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	52	1856	725	1853	928	928	275	1	33	34	50
43	HCECA49	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	53	1558	310	1408	393	393	276	1			1
44	HCEEC15	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	54	948	1	948	9	9	277	1	23	24	65
45	HCESE40	97974 04/04/97 209080 05/29/97	pBluescript	55	990	99	990	193	193	278	1	32	33	256

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO.	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO.	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
45	HCESEF40	97974 04/04/97 209080 05/29/97	pBluescript	224 X	1384	99	1384	193	193	447	1	32	33	205
46	HCFMV39	97974 04/04/97 209080 05/29/97	pSport1	56	1603	1	1296	96	96	279	1	29	30	102
47	HCMSX86	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	57	1052	5	786	12	12	280	1	28	29	32
48	HCNAP62	97975 04/04/97 209081 05/29/97	Lambda ZAP II	58	814	1	558	93	93	281	1	22	23	42
49	HCRAF32	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	59	1215	257	1215		356	282	1	19	20	20
50	HCUDC07	97975 04/04/97 209081 05/29/97	ZAP Express	60	478	1	478	147	147	283	1	36	37	69
51	HCWBB42	97975 04/04/97 209081	ZAP Express	61	618	1	618	212	212	284	1	35	36	74

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
52	HDTAB05	97975 04/04/97 209081	pcMVSpot 2.0	62	751	1	751	257	257	285	1	21	22	32
53	HE2AV74	97975 04/04/97 209081	Uni-ZAP XR	63	780	283	780		433	286	1			16
54	HE2AV71	97975 04/04/97 209081	Uni-ZAP XR	64	588	21	588	169	169	287	1			16
55	HE2GS36	97975 04/04/97 209081	Uni-ZAP XR	65	774	272	774	445	445	288	1			37
56	HE2OF09	97975 04/04/97 209081	Uni-ZAP XR	66	1866	1313	1866	1596	1596	289	1			11
57	HE6EU50	97975 04/04/97 209081	Uni-ZAP XR	67	1152	117	686	237	237	290	1	20	21	34
58	HE9HU17	97975 04/04/97	Uni-ZAP XR	68	2483	1577	2448	1620	1620	291	1			14



Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
59	HE9ND48	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	69	536	1	536	83	83	292	1	36	37	43
60	HEBBW11	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	70	865	647	865		388	293	1	30	31	135
61	HELDY74	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	71	932	1	932	201	201	294	1	17	18	33
62	HEMAE80	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	72	996	1	945	12	12	295	1	24	25	136
63	HFEBA88	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	73	785	464	785	356	356	296	1	29	30	57
64	HFGAB89	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	74	1069	196	1047	295	295	297	1	32	33	34
65	HFEVHY45	97975	pBluescript	75	831	1	831		89	298	1	30	31	76

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
66	HGBA193	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	76	590	1	590	233	233	299	1	38	39	94
67	HGBBQ69	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	77	1274	1	1273	105	105	300	1	24	25	43
68	HHFCH08	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	78	1133	4	1042	175	175	301	1	23	24	30
69	HHFHJ59	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	79	661	1	661	192	192	302	1	29	30	112
70	HHFHR32	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	80	1378	1	1378		358	303	1			13
71	HHGCN69	97975 04/04/97 209081 05/29/97	Lambda ZAP II	81	1440	298	1440	532	532	304	1	23	24	34

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig. Pep	Last AA of Sig. Pep	First AA of Secreted Portion	Last AA of ORF
72	HHGDO13	97975 04/04/97 209081 05/29/97	Lambda ZAP II	82	1381	766	1371	993	993	305	1	23	24	34
73	HHPFED63	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	83	1706	182	1644	257	257	306	1	24	25	81
74	HHSEG23	97976 04/04/97	Uni-ZAP XR	84	573	1	573	160	160	307	1	18	19	71
75	HJPAV06	97976 04/04/97	Uni-ZAP XR	85	684	199	684	323	323	308	1	27	28	33
76	HKIXL73	97976 04/04/97	pBluescript	86	1036	591	1036	690	690	309	1	32	33	114
77	HKMNC43	97976 04/04/97	pBluescript	87	908	1	908	139	139	310	1	18	19	108
78	HMEJE31	97976 04/04/97	Lambda ZAP II	88	655	1	655	165	165	311	1	33	34	64
79	HMSKS35	97976 04/04/97	Uni-ZAP XR	89	1102	1	1102	228	228	312	1	26	27	49
80	HNF AE54	97976 04/04/97	Uni-ZAP XR	90	1533	665	1518	347	347	313	1	26	27	293
81	HNFJH45	97976 04/04/97	Uni-ZAP XR	91	575	1	575	275	275	314	1	30	31	67
82	HNGBT31	97976 04/04/97	Uni-ZAP XR	92	639	1	639	224	224	315	1	28	29	104

Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
83	HNGIN60	97976 04/04/97	Uni-ZAP XR	93	744	1	744	225	225	316	1	43	44	70
84	HNGJG84	97976 04/04/97	Uni-ZAP XR	94	526	1	526	268	268	317	1	29	30	38
85	HNHDW42	97976 04/04/97	Uni-ZAP XR	95	426	1	426	168	168	318	1	28	29	71
86	HNHFL57	97976 04/04/97	Uni-ZAP XR	96	844	1	844	98	98	319	1	25	26	61
87	HOGAR52	97977 04/04/97 209082 05/29/97	pcMVSPORT 2.0	97	1985	453	1985	533	533	320	1	17	18	285
88	HOSBZ55	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	98	1416	69	1416	246	246	321	1	32	33	54
89	HOSDI92	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	99	1935	141	772		274	322	1	20	21	58
90	HPBCU51	97977 04/04/97 209082 05/29/97	pBluescript SK-	100	599	1	599	86	86	323	1	27	28	119
91	HPCAL49	97977 04/04/97 209082	Uni-ZAP XR	101	784	1	784		280	324	1	18	19	43

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
92	HPFCR13	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	102	1035	602	1035	859	859	325	1	32	33	58
93	HPHAC83	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	103	2218	840	2182	1035	1035	326	1	17	18	17
94	HPMBQ32	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	104	1351	1	1351	18	18	327	1	23	24	86
95	HPWAN23	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	105	2066	51	2052	270	270	328	1	29	30	537
95	HPWAN23	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	226	2057	1	1954	220	220	449	1	29	30	315
96	HRDFB85	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	106	1705	23	1697	233	233	329	1	21	22	201
97	HRGBR28	97977 04/04/97	Uni-ZAP XR	107	1167	611	1167	53	53	330	1	1	2	263

Gene No.	cDNA Clone ID	ATCC Deposit Nr. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		209082 05/29/97												
98	HSKGN81	97977 04/04/97 209082 05/29/97	pBluescript	108	1907	151	1432	353	353	331	1	23	24	260
98	HSKGN81	97977 04/04/97 209082 05/29/97	pBluescript	227	2084	335	2084	537	537	450	1	19	20	23
99	HSPA56	97977 04/04/97 209082 05/29/97	pSport1	109	611	1	576	229	229	332	1	25	26	47
100	HE8EU04	209746 04/07/98	Uni-ZAP XR	110	2632	294	2632	337	337	333	1	25	26	333
100	HSXBT86	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	228	2143	53	1096	235	235	451	1			9
101	HSXCS62	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	111	2249	1	1953	90	90	334	1	18	19	199
102	HTIEU09	97977 04/04/97 209082	Uni-ZAP XR	112	2198	228	2158	400	400	335	1			23

Gene No.	cDNA Clone ID	ATCC Deposit Nr. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
103	HTTEKM35	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	113	1043	40	1043	320	320	336	1	20	21	142
104	HTGEP89	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	114	703	1	703	285	285	337	1	29	30	94
105	HTGEW91	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	115	3684	526	1338	584	584	338	1	24	25	37
106	HTOEY16	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	116	1965	127	1915	202	202	339	1	27	28	38
107	HTPCN79	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	117	503	1	503		1	340	1	7	8	70
108	HTSGM54	97977 04/04/97 209082 05/29/97	pBluescript	118	1133	316	1069		423	341	1	12	13	84
109	HTSHE40	97977 04/04/97	pBluescript	119	1101	118	956	218	218	342	1	31	32	89

Gene No.	cDNA Clone ID	ATCC Deposit, Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		209082 05/29/97												
110	HTWAF58	97977 04/04/97 209082 05/29/97	Lambda ZAP II	120	282	1	282	137	137	343	1	25	26	48
111	HTWBY29	97977 04/04/97 209082 05/29/97	pSport1	121	2635	1593	2489	1654	1654	344	1	25	26	55
112	HUKFC71	209007 04/28/97 209083 05/29/97	Lambda ZAP II	122	994	1	932		272	345	1	15	16	221
113	HCE3Q10	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	123	1542	1	1542	143	143	346	1	25	26	63
114	HCEVR60	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	124	1390	82	1390	127	127	347	1	32	33	153
115	HDTAW95	209007 04/28/97 209083 05/29/97	pcMVSpport 2.0	125	1288	412	1288	571	571	348	1			16
116	HE6EL90	209007	Uni-ZAP XR	126	1517	1	1452	243	243	349	1			9





Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
123	HOABL56	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	133	1720	565	1720	660	660	356	1	18	19	21
124	HPMCJ92	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	134	705	28	705	106	106	357	1	28	29	98
125	HPWAZ95	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	135	323	1	323	88	88	358	1	27	28	78
126	HRGBR18	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	136	582	1	582		16	359	1	17	18	30
127	HSUBW09	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	137	1021	1	1021	153	153	360	1	32	33	56
128	HUKCO64	209007 04/28/97 209083 05/29/97	Lambda ZAP II	138	1777	439	1777		521	361	1			2
129	H6EAA53	209007 04/28/97 209083	Uni-ZAP XR	139	643	303	643		313	362	1	7	8	31

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
130	HAGA111	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	140	1220	1	1220		127	363	1	16	17	27
131	HAGA039	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	141	721	1	721		415	364	1			14
132	HALSK07	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	142	1468	125	1468	210	210	365	1	29	30	33
133	HALSQ59	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	143	300	4	300	101	101	366	1	22	23	66
134	HAIBP89	unknown 05/18/98	Uni-ZAP XR	144	2243	173	2243	311	311	367	1	27	28	317
134	HGCB91	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	229	1025	409	1025	624	624	452	1	20	21	25
135	HBMTD81	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	145	1082	163	1082	357	357	368	1			30

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
136	HBXGK12	209008 04/28/97 209084 05/29/97	ZAP Express	146	4313	1153	4313	1313	1313	369	1	18	19	42
137	HFKFJ07	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	147	1183	1	1183	149	149	370	1	41	42	254
138	HCOA140	209008 04/28/97 209084 05/29/97	Lambda ZAP II	148	734	1	734	285	285	371	1			19
139	HCWHZ24	209008 04/28/97 209084 05/29/97	ZAP Express	149	1405	1	1405	108	108	372	1	34	35	63
140	HE2GT20	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	150	2890	1178	2890	1178	1178	373	1	31	32	39
141	HE8EY43	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	151	2399	1181	2399	1265	1265	374	1	30	31	34
142	HFCB37	209008 04/28/97 209084	Uni-ZAP XR	152	802	352	802		487	375	1			10

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
143	HFTCT67	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	153	461	24	461	145	145	376	1	37	38	63
144	HGLAM46	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	154	2388	818	2388	648	648	377	1			18
145	HHGBR15	209008 04/28/97 209084 05/29/97	Lambda ZAP II	155	642	322	642	400	400	378	1			4
146	HJAAU36	209008 04/28/97 209084 05/29/97	pBluescript SK-	156	1251	583	1251		933	379	1	16	17	16
147	HUSIT49	209008 04/28/97 209084 05/29/97	pSport1	157	2127	247	2127	383	383	380	1	47	48	83
148	HKLAB16	209008 04/28/97 209084 05/29/97	Lambda ZAP II	158	1625	817	1625	1012	1012	381	1	18	19	20
149	HLMMU76	209008 04/28/97	Lambda ZAP II	159	1687	1307	1687	1296	1296	382	1	28	29	28

Gene No.	cDNA Clone ID	ATCC Deposit Nr. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		209084 05/29/97												
150	HMSKQ35	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	160	1842	172	1463	319	319	383	1	30	31	33
151	HNHED86	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	161	770	1	770	30	30	384	1	31	32	46
152	HNHEJ88	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	162	519	1	519	242	242	385	1	17	18	24
153	HNHFQ63	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	163	753	1	753	164	164	386	1	17	18	67
154	HOECU83	209009 04/28/97	Uni-ZAP XR	164	1400	189	1400		508	387	1	22	23	33
155	HPTRC15	209009 04/28/97	pBluescript	165	2153	594	2153		611	388	1			13
156	HSKCP69	209009 04/28/97	Uni-ZAP XR	166	1251	219	1120			389	1			
156	HSKCP69	209009 04/28/97	Uni-ZAP XR	230	1250	223	1250	393	393	453	1	32	33	171
157	H6EAE26	209009	Uni-ZAP XR	167	882	48	882	155	155	390	1	33	34	153

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
158	HAGBX03	209009 04/28/97	Uni-ZAP XR	168	1208	1	1208	182	182	391	1			8
159	HAGDQ47	209009 04/28/97	Uni-ZAP XR	169	1307	1	1307	44	44	392	1	22	23	60
160	HAICP19	209009 04/28/97	Uni-ZAP XR	170	1624	89	1483	128	128	393	1	18	19	446
161	HAUAE83	209009 04/28/97	Uni-ZAP XR	171	2003	889	2003	1080	1080	394	1			23
162	HBHAD12	209009 04/28/97	Uni-ZAP XR	172	786	1	786		176	395	1	17	18	23
163	HBMITY28	209009 04/28/97	Uni-ZAP XR	173	1758	962	1758	1184	1184	396	1	27	28	34
164	HBMVTP04	209009 04/28/97	Uni-ZAP XR	174	888	330	862		546	397	1			2
165	HCDDDB78	209009 04/28/97	Uni-ZAP XR	175	2379	750	2379	901	901	398	1	18	19	24
166	HCEQA68	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	176	1348	1	1348	12	12	399	1	28	29	78
167	HCEZS40	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	177	1502	178	1502	315	315	400	1			20
168	HCFNF11	209010	pSport1	178	1637	26	1607	152	152	401	1	44	45	257

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		04/28/97 209085 05/29/97												
169	HCRBL20	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	179	2911	1103	2858	192	192	402	1	32	33	424
169	HCRBL20	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	231	1811	20	1811	93	93	454	1	36	37	95
170	HCUBL62	209010 04/28/97 209085 05/29/97	ZAP Express	180	519	1	519	57	57	403	1	28	29	32
171	HDSAP81	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	181	968	320	968	476	476	404	1	27	28	79
172	HE2CT29	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	182	1128	1	1128	111	111	405	1	26	27	94
173	HE8MG65	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	183	2276	48	2276	88	88	406	1	37	38	257



Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
173	HE8MG65	209010 04/28/97 209085	Uni-ZAP XR	232	2271	56	2232	79	79	455	1	43	44	170
174	HE9FB42	209010 04/28/97 209085	Uni-ZAP XR	184	2500	76	1693	518	518	407	1	1	2	623
175	HEMAM41	209010 04/28/97 209085	Uni-ZAP XR	185	1337	60	1328	175	175	408	1	39	40	190
175	HEMAM41	209010 04/28/97 209085	Uni-ZAP XR	233	1338	33	1327	175	175	456	1	32	33	91
176	HEMCV19	209010 04/28/97 209085	Uni-ZAP XR	186	941	33	931	79	79	409	1	23	24	178
177	HEMDX17	209010 04/28/97 209085	Uni-ZAP XR	187	654	1	654	137	137	410	1			13
178	HEIAR54	209010 04/28/97 209085	Uni-ZAP XR	188	1848	454	1848	948	948	411	1	14	15	232

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
179	HETBX14	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	189	1146	157	1146		74	412	1	14	15	53
180	HFGAB48	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	190	906	156	906	245	245	413	1	30	31	32
181	HFKF140	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	191	1941	120	1002	213	213	414	1	18	19	218
182	HFXHN68	209010 04/28/97 209085 05/29/97	Lambda ZAP II	192	2118	777	2118	966	966	415	1	23	24	50
183	HGBFO79	209011 04/28/97	Uni-ZAP XR	193	1538	259	1538	273	273	416	1	23	24	49
184	HGLAM56	209011 04/28/97	Uni-ZAP XR	194	1098	68	1098		185	417	1	28	29	69
185	HHLBA89	209011 04/28/97	pBluescript SK-	195	1001	1	1001	324	324	418	1	25	26	39
186	HHPDW05	209011 04/28/97	Uni-ZAP XR	196	1443	1	1443	246	246	419	1	21	22	21
187	HIPSD37	209011 04/28/97	pBluescript	197	1282	66	1282	171	171	420	1	19	20	37

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
188	HHPSF70	209011 04/28/97	pBluescript	198	951	26	951		162	421	1	16	17	34
189	HHSAK25	209011 04/28/97	Uni-ZAP XR	199	1740	1390	1740	1534	1534	422	1	19	20	31
190	HIASB53	209011 04/28/97	pBluescript	200	1707	401	1195	652	652	423	1	26	27	126
191	HIABZ65	209011 04/28/97	pBluescript SK-	201	779	1	779	23	23	424	1	26	27	68
192	HIPIB39	209011 04/28/97	Uni-ZAP XR	202	1617	188	1605	182	182	425	1	28	29	91
193	HLHSK94	209011 04/28/97	pBluescript	203	1974	1	1794	112	112	426	1	26	27	379
194	HLHTC70	209011 04/28/97	pBluescript	204	1057	229	1057	365	365	427	1	23	24	22
195	HLMIW92	209011 04/28/97	Lambda ZAP II	205	721	1	721	244	244	428	1	25	26	46
196	HLTCY93	209011 04/28/97	Uni-ZAP XR	206	2465	988	2465	1225	1225	429	1			4
197	HLTDB65	209011 04/28/97	Uni-ZAP XR	207	1480	1	1480		371	430	1	15	16	143
198	HMSHM43	209011 04/28/97	Uni-ZAP XR	208	872	1	872	35	35	431	1	18	19	36
199	HMSHQ24	209011 04/28/97	Uni-ZAP XR	209	1779	16	1779	148	148	432	1	24	25	36
200	HNF4H08	209011 04/28/97	Uni-ZAP XR	210	2110	592	2110	611	611	433	1	18	19	191

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
201	HNGAO10	209011 04/28/97	Uni-ZAP XR	211	938	1	938	107	107	434	1	27	28	30
202	HNGBE45	209011 04/28/97	Uni-ZAP XR	212	1551	1	1551	114	114	435	1	21	22	100
203	HNHAZ16	209011 04/28/97	Uni-ZAP XR	213	997	1	997	202	202	436	1	24	25	36
204	HNHCM59	209011 04/28/97	Uni-ZAP XR	214	1496	1	1132		165	437	1	28	29	41
205	HOSFM22	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	215	1308	501	1308		809	438	1			1
206	HPHAC88	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	216	1705	384	1705	549	549	439	1	23	24	24
207	HCD EO95	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	217	999	608	999	273	273	440	1	22	23	54

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

5       The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, *Gene* 67:31-40 (1988).

Polypeptides of the invention also can be purified from natural or recombinant sources  
10       using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

### Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the  
15       cleavage point for that sequence, are available. For instance, the method of McGeoch, *Virus Res.* 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, *Nucleic Acids Res.* 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1  
20       indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, *supra.*) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide  
25       was analyzed by a computer program called SignalP (Henrik Nielsen et al., *Protein Engineering* 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results  
30       shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., +  
35       or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

#### 10 Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

15 By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown in Table 1, the ORF (open reading frame), or any fragment specified as described herein.

25 As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are:

35 Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization



Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query

amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity.

For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level.

Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after

deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as to have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

### Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the

carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Particularly, N-terminal deletions of the polypeptide of the present invention can be described by the general formula m-p, where p is the total number of amino acids in the polypeptide and m is an integer from 2 to (p-1), and where both of these integers (m & p) correspond to the position of the amino acid residue identified in SEQ ID NO:Y.

Moreover, C-terminal deletions of the polypeptide of the present invention can also be described by the general formula 1-n, where n is an integer from 2 to (p-1), and again where these integers (n & p) correspond to the position of the amino acid residue identified in SEQ ID NO:Y.

The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini, which may be described generally as having residues m-n of SEQ ID NO:Y, where m and n are integers as described above.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

### Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an

epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')<sub>2</sub> fragments) which are capable of specifically binding to protein. Fab and F(ab')<sub>2</sub> fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

### Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the



polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

5        Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

10        Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final  
15        preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

20        Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG)  
25        can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

30        Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the  
35        fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D.

Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein.

Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

#### Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance

genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and psrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., *Basic Methods In Molecular Biology* (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein.

after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

5

### Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

10

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

15

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

20

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

25

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

30

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are

35

more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991) ) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model

systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of

unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

### Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine ( $^{125}\text{I}$ ,  $^{121}\text{I}$ ), carbon ( $^{14}\text{C}$ ), sulfur ( $^{35}\text{S}$ ), tritium ( $^3\text{H}$ ), indium ( $^{112}\text{In}$ ), and technetium ( $^{99\text{m}}\text{Tc}$ ), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example,  $^{131}\text{I}$ ,  $^{112}\text{In}$ ,  $^{99\text{m}}\text{Tc}$ ), a radio-opaque substance, or a material detectable by nuclear magnetic

resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human  
5 subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of <sup>99m</sup>Tc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The  
10 Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene  
15 expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to  
20 supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired  
25 response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such  
30 as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a  
35 recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.



### Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

### Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia); or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can

decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic

shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

### Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenström's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

### Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases

may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

5        Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, 10    Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., 15    Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, 20    Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

      Similarly, bacterial or fungal agents that can cause disease or symptoms and that 25    can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, 30    Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Meningococcal), Pasteurellaceae Infections (e.g., Actinobacillus, Haemophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, 35    and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS

related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria.

5 Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

10 Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas.

15 These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or

20 diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo

25 therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

### Regeneration

A polynucleotide or polypeptide of the present invention can be used to

30 differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion

35 injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal

or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase  
5 regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue  
10 regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate  
15 nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease,  
20 Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

### Chemotaxis

25 A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular  
30 trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to  
35 treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

5

### Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit  
10 (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural  
15 or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell  
20 membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing  
25 the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a  
30 labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product  
mixtures. The assay may also simply comprise the steps of mixing a candidate  
35 compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

5 All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

10 Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with  
15 a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

#### Other Activities

20 A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic  
25 surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, circadian  
30 rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

35 A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.



### Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

5 Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

10 Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

15 Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

20 Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

25 A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

30 A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

35 A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method

comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

5 Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined  
10 from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95%  
15 identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

20 The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

25 Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous  
30 nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

35 The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95%

identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is  
5 at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the  
10 deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

15 Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

20 Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the  
25 amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the  
30 complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein  
35 encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide

comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is  
5 any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

10 Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid  
15 sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising  
20 inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising  
25 culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID  
30 NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the  
35 deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

### Examples

#### Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector.

Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited Plasmid</u>
Lambda Zap	pBluescript (pBS)
Uni-Zap XR	pBluescript (pBS)
Zap Express	pBK
lafmid BA	plafmid BA
pSport1	pSport1
pCMVSPORT 2.0	pCMVSPORT 2.0
pCMVSPORT 3.0	pCMVSPORT 3.0
pCR <sup>®</sup> 2.1	pCR <sup>®</sup> 2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Altting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Altting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1



Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS-. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the fl origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the fl ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into *E. coli* strain XL-1 Blue: Vector pCR<sup>2.1</sup>, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, available from Life Technologies: (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with <sup>32</sup>P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).)

The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance; in 25  $\mu$ l of reaction mixture with 0.5  $\mu$ g of the above cDNA template. A convenient reaction mixture is 1.5-5 mM  $MgCl_2$ , 0.01% (w/v) gelatin, 20  $\mu$ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., *Nucleic Acids Res.* 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then  
5 be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA  
10 synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

15  
**Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide**

A human genomic PI library (Genomic Systems, Inc.) is screened by PCR  
20 using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

**Example 3: Tissue Distribution of Polypeptide**

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by,  
25 among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P<sup>32</sup> using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is  
30 then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are  
35 mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

**Example 4: Chromosomal Mapping of the Polynucleotides**

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

**Example 5: Bacterial Expression of a Polypeptide**

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp<sup>r</sup>), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan<sup>r</sup>). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.<sup>600</sup>) of between 0.4 and 0.6. IPTG

(Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1-5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA

insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

- 5           The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

**Example 6: Purification of a Polypeptide from an Inclusion Body**

- 10           The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

- Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

- 20           The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

- 25           The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

- 30           Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

          To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area

(e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant  $A_{280}$  monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commaissie blue stained 16% SDS-PAGE gel when 5  $\mu$ g of purified protein is loaded.

The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

### Example 7: Cloning and Expression of a Polypeptide in a Baculovirus

#### Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak *Drosophila* promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five  $\mu$ g of a plasmid containing the polynucleotide is co-transfected with 1.0  $\mu$ g of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417 (1987). One  $\mu$ g of BaculoGold™ virus DNA and 5  $\mu$ g of the plasmid are mixed in a sterile well of a microtiter plate containing 50  $\mu$ l of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10  $\mu$ l Lipofectin plus 90  $\mu$ l Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm



tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

- 5 After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.)
- 10 After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in
- 15 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

- To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection
- 20 ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of <sup>35</sup>S-methionine and 5 µCi <sup>35</sup>S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins
- 25 in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

### 30 Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates

the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV1, HIV1 and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., *J. Biol. Chem.* 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., *Biochem. et Biophys. Acta*, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., *Biotechnology* 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., *Biochem J.* 227:277-279 (1991); Bebbington et al., *Bio/Technology* 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., *Molecular and Cellular Biology*, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., *Cell* 41:521-530 (1985)). Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the

polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five  $\mu$ g of the expression plasmid pC6 is cotransfected with 0.5  $\mu$ g of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1  $\mu$ M, 2  $\mu$ M, 5  $\mu$ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200  $\mu$ M. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

**Example 9: Protein Fusions**

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the half-life time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

```

GGGATCCGGAGCCCAAATCTTCTGACAAACTCACACATGCCCACCGTGCC
CAGCACCTGAATTTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCCAAACC
35 CAAGGACACCCTCATGATCTCCCGGACTCCTGAGGTACATGCGTGGTGGT
GGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACG
GCGTGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC

```

AGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTG  
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAACCCCC  
ATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGT  
GTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCT  
5 GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGGA  
GAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGG  
ACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCA  
GGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGC  
ACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGC  
10 GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

**Example 10: Production of an Antibody from a Polypeptide**

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

20 In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at 30 about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line 35 (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as

described by Wands et al. (Gastroenterology 80:225-232 (1981)). The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')<sub>2</sub> and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')<sub>2</sub> fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

#### **Example 11: Production Of Secreted Protein For High-Throughput Screening Assays**

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20:

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a

working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The  
 5 PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at  $2 \times 10^5$  cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x

10 Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in

15 Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of  
 20 transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off  
 25 PBS rinse, and person-B, using a 12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of  $\text{CaCl}_2$  (anhyd); 0.00130 mg/L  
 30  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ ; 0.050 mg/L of  $\text{Fe}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$ ; 0.417 mg/L of  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ ; 311.80 mg/L of KCl; 28.64 mg/L of  $\text{MgCl}_2$ ; 48.84 mg/L of  $\text{MgSO}_4$ ; 6995.50 mg/L of NaCl; 2400.0 mg/L of  $\text{NaHCO}_3$ ; 62.50 mg/L of  $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ ; 71.02 mg/L of  $\text{Na}_2\text{HPO}_4$ ; .4320 mg/L of  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ ; .002 mg/L of Arachidonic Acid ; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic  
 35 Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitic Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of

Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L-Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H<sub>2</sub>O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H<sub>2</sub>O; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H<sub>2</sub>O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalanine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tyrosine-2Na-2H<sub>2</sub>O; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B<sub>12</sub>; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine; 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mM glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.



### Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

	<u>Ligand</u>	<u>tyk2</u>	<u>JAKs</u>			<u>STATs</u>	<u>GAS(elements) or ISRE</u>
			<u>Jak1</u>	<u>Jak2</u>	<u>Jak3</u>		
	<u>IFN family</u>						
5	IFN- $\alpha$ /B	+	+	-	-	1,2,3	ISRE
	IFN-g		+	+	-	1	GAS (IRF1>Lys6>IFP)
	IL-10	+	?	?	-	1,3	
	<u>gp130 family</u>						
10	IL-6 (Pleiotrohic)	+	+	+	?	1,3	GAS (IRF1>Lys6>IFP)
	IL-11(Pleiotrohic)	?	+	?	?	1,3	
	OnM(Pleiotrohic)	?	+	+	?	1,3	
	LIF(Pleiotrohic)	?	+	+	?	1,3	
	CNTF(Pleiotrohic)	-/+	+	+	?	1,3	
15	G-CSF(Pleiotrohic)	?	+	?	?	1,3	
	IL-12(Pleiotrohic)	+	-	+	+	1,3	
	<u>g-C family</u>						
20	IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
	IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >>Ly6)(IgH)
	IL-7 (lymphocytes)	-	+	-	+	5	GAS
	IL-9 (lymphocytes)	-	+	-	+	5	GAS
	IL-13 (lymphocyte)	-	+	?	?	6	GAS
25	IL-15	?	+	?	+	5	GAS
	<u>gp140 family</u>						
	IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)
	IL-5 (myeloid)	-	-	+	-	5	GAS
30	GM-CSF (myeloid)	-	-	+	-	5	GAS
	<u>Growth hormone family</u>						
	GH	?	-	+	-	5	
	PRL	?	+/-	+	-	1,3,5	
35	EPO	?	-	+	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)
	<u>Receptor Tyrosine Kinases</u>						
	EGF	?	+	+	-	1,3	GAS (IRF1)
	PDGF	?	+	+	-	1,3	
40	CSF-1	?	+	+	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5':GCGCCTCGAGATTTCCTCCGAAATCTAGATTTCCTCCGAAATGATTTCCTCCG  
AAATGATTTCCTCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTGGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5':CTCGAGATTTCCTCCGAAATCTAGATTTCCTCCGAAATGATTTCCTCCGAAATG  
ATTTTCCTCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC  
CTAACTCCGCCCCTCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGC  
CCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCGGC  
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTT  
TGCAAAAAGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

**Example 13: High-Throughput Screening Assay for T-cell Activity.**

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1% Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells ( $10^7$  per transfection), and resuspend in OPTI-MEM to a final concentration of  $10^7$  cells/ml. Then add 1ml of  $1 \times 10^7$  cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentamicin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophane covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

**Example 14: High-Throughput Screening Assay Identifying Myeloid Activity**

The following protocol is used to assess myeloid activity-by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells.

5 Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line; although TF-1, HL60, or KG1 can be used.

10 To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest  $2 \times 10^7$  U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

15 Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 1 mM  $\text{MgCl}_2$ , and 675 uM  $\text{CaCl}_2$ . Incubate at 37°C for 45 min.

20 Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

25 These cells are tested by harvesting  $1 \times 10^8$  cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of  $5 \times 10^5$  cells/ml. Plate 200 ul cells per well in the 96-well plate (or  $1 \times 10^5$  cells/well).

30 Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat pheochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine

growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as  $5 \times 10^5$  cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to  $1 \times 10^5$  cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

#### Example 16: High-Throughput Screening Assay for T-cell Activity

NF- $\kappa$ B (Nuclear Factor  $\kappa$ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- $\kappa$ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- $\kappa$ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- $\kappa$ B is retained in the cytoplasm with I- $\kappa$ B (Inhibitor  $\kappa$ B). However, upon stimulation, I- $\kappa$ B is phosphorylated and degraded, causing NF- $\kappa$ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- $\kappa$ B include IL-2, IL-6, GM-CSF, ICAM-1 and class I MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- $\kappa$ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- $\kappa$ B would be useful in treating



diseases. For example, inhibitors of NF- $\kappa$ B could be used to treat those diseases related to the acute or chronic activation of NF- $\kappa$ B, such as rheumatoid arthritis.

To construct a vector containing the NF- $\kappa$ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- $\kappa$ B binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:  
5':GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGAC  
TTTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:  
5':GCGGCAAGCTTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene)

Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGGACTTTCC  
ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCA  
TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTTCTCCGCCCCATGGCTGACT  
AATTTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCGGCCTCTGAGCTATTC  
CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTT:  
3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2- promoter plasmid (Clontech) with this NF- $\kappa$ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- $\kappa$ B/SV40/SEAP cassette is removed from the above NF- $\kappa$ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the NF- $\kappa$ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF- $\kappa$ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

#### Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15  $\mu$ l of 2.5x dilution buffer into Optiplates containing 35  $\mu$ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50  $\mu$ l Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50  $\mu$ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

#### Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6

23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

**Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability**

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO<sub>2</sub> incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50  $\mu$ l of 12  $\mu$ g/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO<sub>2</sub> incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100  $\mu$ l of buffer.

5 For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to  $2.5 \times 10^6$  cells/ml with HBSS in a 50-ml conical tube. 4  $\mu$ l of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to  $1 \times 10^6$  cells/ml, and dispensed into a microplate, 100  
10  $\mu$ l/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200  $\mu$ l, followed by an aspiration step to 100  $\mu$ l final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the  
15 following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50  $\mu$ l. Increased emission at 530-nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca<sup>++</sup> concentration.

20

#### Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase  
25 RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

30 Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members  
35 of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na<sub>3</sub>VO<sub>4</sub>, 2 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub> and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a

biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

- 5        The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg<sub>2+</sub> (5mM ATP/50mM MgCl<sub>2</sub>), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl<sub>2</sub>, 5 mM MnCl<sub>2</sub>, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the
- 10       components gently and preincubate the reaction mix at 30°C for 2 min. Initiate the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

- 15       Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as
- 20       above.

- Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of
- 25       tyrosine kinase activity.

#### **Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity**

- 30       As a potential alternative and/or complement to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase,
- 35       Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other

phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFLA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

#### Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated.. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies).

The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., *Nucleic Acids Research*, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenin deoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., *Methods Cell Biol.* 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cy. et al., *Genet. Anal. Tech. Appl.*, 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

#### **Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.



The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

#### **Example 23: Formulating a Polypeptide**

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1  $\mu\text{g/kg/day}$  to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1  $\mu\text{g/kg/hour}$  to about 50  $\mu\text{g/kg/hour}$ , either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch); buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules.

Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman; U. et al., Biopolymers 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid; and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

**Example 24: Method of Treating Decreased Levels of the Polypeptide**

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form.

5 Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

10 For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

**Example 25: Method of Treating Increased Levels of the Polypeptide**

15 Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

**Example 26: Method of Treatment Using Gene Therapy**

25 One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is 30 turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 27: Method of Treatment Using Gene Therapy - In Vivo

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide of the present invention. A polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the encoded polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, WO90/11092, WO98/11779; U.S. Patent NO. 5693622, 5705151, 5580859; Tabata H. et al. (1997) Cardiovasc. Res. 35(3):470-479, Chao J et al. (1997) Pharmacol. Res. 35(6):517-522, Wolff J.A. (1997) Neuromuscul. Disord. 7(5):314-318, Schwartz B. et al. (1996) Gene Ther. 3(5):405-411, Tsurumi Y. et al. (1996) Circulation 94(12):3281-3290 (incorporated herein by reference).

The polynucleotide constructs of the present invention may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). These polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides may also be delivered in liposome formulations (such as those taught in Felgner P.L. et al. (1995) Ann. NY Acad. Sci. 772:126-139 and Abdallah B. et al. (1995) Biol. Cell 85(1):1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs of the present invention used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct of the present invention can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 g/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for the polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with

liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice. The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA of the present invention.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.



*Sequence Listing*

## (1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Human Genome Sciences, Inc., et al.
- (ii) TITLE OF INVENTION: 207 Human Secreted Proteins
- 10 (iii) NUMBER OF SEQUENCES: 800
- (iv) CORRESPONDENCE ADDRESS:
- 15 (A) ADDRESSEE: Human Genome Sciences, Inc.
- (B) STREET: 9410 Key West Avenue
- 20 (C) CITY: Rockville
- (D) STATE: Maryland
- (E) COUNTRY: USA
- 25 (F) ZIP: 20850
- (v) COMPUTER READABLE FORM:
- 30 (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
- (B) COMPUTER: HP Vectra 486/33
- 35 (C) OPERATING SYSTEM: MSDOS version 6.2
- (D) SOFTWARE: ASCII Text
- 40 (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
- 45 (B) FILING DATE:
- (C) CLASSIFICATION:
- 50 (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER:
- 55 (B) FILING DATE:

## (viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Kenley K. Hoover  
(B) REGISTRATION NUMBER: 40,302  
(C) REFERENCE/DOCKET NUMBER: P2007PCT

## (vi) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (301) 309-8504  
(B) TELEFAX: (301) 309-8439

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 733 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGGATCCGGA GCCCAAATCT TCTGACAAAA CTCACACATG CCCACCGTGC CCAGCACCTG 60  
AATTCGAGGG TGCACCGTCA GTCTTCCTCT TCCCCCAAA ACCCAAGGAC ACCCTCATGA 120  
TCTCCCGGAC TCCTGAGGTC ACATGCGTGG TGGTGGACGT AAGCCACGAA GACCCTGAGG 180  
TCAAGTTCAA CTGGTACGTG GACGGCGTGG AGGTGCATAA TGCCAAGACA AAGCCGCGGG 240  
AGGAGCAGTA CAACAGCAGG TACCGTGTGG TCAGCGTCCT CACCGTCCTG CACCAGGACT 300  
GGCTGAATGG CAAGGAGTAC AAGTGCAAGG TCTCCAACAA AGCCCTCCCA ACCCCCATCG 360  
AGAAAACCAT CTCCAAGCC AAAGGGCAGC CCGGAGAACC ACAGGTGTAC ACCCTGCCCC 420  
CATCCCGGGA TGAGCTGACC AAGAACCAGG TCAGCCTGAC CTGCCTGGTC AAAGGCTTCT 480  
ATCCAAGCGA CATCGCCGTG GAGTGGGAGA GCAATGGGCA GCCGGAGAAC AACTACAAGA 540  
CCACGCCTCC CGTGCTGGAC TCCGACGGCT CTTTCTTCT CTACAGCAAG CTCACCGTGG 600  
ACAAGAGCAG GTGGCAGCAG GGAACGTCT TCTCATGCTC CGTGATGCAT GAGGCTCTGC 660  
ACAACCACTA CACGCAGAAG AGCCTCTCCC TGTCTCCGGG TAAATGAGTG CGACGGCCGC 720  
GACTCTAGAG GAT 733

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Ser Xaa Trp Ser  
1 5

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 86 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GCGCCTCGAG ATTTCCTCGA AATCTAGATT TCCCCGAAAT GATTTCCTCG AAATGATTTC 60  
CCCGAAATAT CTGCCATCTC AATTAG 86

## (2) INFORMATION FOR SEQ ID NO: 4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GCGGCAAGCT TTTTGCAAAG CCTAGGC 27

## (2) INFORMATION FOR SEQ ID NO: 5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 271 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGATTT CCCCCGAATC TAGATTTCCT CGAAATGATT TCCCCGAAAT GATTTCCTCG 60  
AAATATCTGC CATCTCAATT AGTCAGCAAC CATAGTCTCG CCCCTAACTC CGCCCATCTC 120

GGCCCTAACT CCGCCGAGTT CCGCCCATTC TCGCCCCAT GGTGACTAA TTTTITTTAT 180  
TTATGCAGAG GCCGAGGCCG CCTCGGCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCTT 240  
5 TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT T 271

10 (2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs  
(B) TYPE: nucleic acid  
15 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

20 GCGCTCGAGG GATGACAGCG ATAGAACCCC GG 32

25 (2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
30 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

35 GCGAAGCTTC GCGACTCCCC GGATCCGCCT C 31

40 (2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs  
45 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

50 GGGGACTTTC CC 12

55 (2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 73 base pairs  
60 (B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

5 GCGGCGCTCGA GGGGACTTTC CCGGGGACTT TCCGGGGACT TTCCGGGACT TTCCATCCTG 60  
CCATCTCAAT TAG 73

10

(2) INFORMATION FOR SEQ ID NO: 10:

15 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 256 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTCGAGGGGA CTTCCTCCGG GACTTTCGG GGACTTTCGG GGACTTTCCA TCTGCCATCT 60  
25 CAATTAGTCA GCAACCATAG TCCCGCCCTT AACTCCGCC ATCCCGCCCC TAACTCCGCC 120  
CAGTTCGGCC CATTCTCGGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA 180  
GGCGGCTCG GCCTCTGAGC TATTCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG 240  
30 CTTTTCGAAA AAGCTT 256

35

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2526 base pairs  
40 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

45 GACAGGCTAT CCGAGAATCT GAGAGCTGGG CCCGGCAATT CCTCCAGTA CCCTTGTGAC 60  
CTAAGTCCAG TCACACATTT CCCAAAGTTT CTCTTTGTCA TAACCCTGGT CTGGCTGGTT 120  
50 TTGRGGRCTT GAGAATGGT CAGGGACTCC AGGCCAAGTC CAACAGAGAC CCCAAACCCA 180  
CCACACACCA GCACCCACAA CCTCACCACC AAGAAAGAGG ACTTTTGTGG GGCCACAAGT 240  
AAGAGGTCAT TTCTGGAATG GACTCAGACC TTAAACAGG AGAGTTGAGC ACTTCCAGKS 300  
55 AGTTTTTAAG CAAGGCATGG GGAACAGGA ATAGAACCTT TCAAAGAGGT TGCCGAGAGA 360  
AAAGCTGGGC CTCTTGCAAT CGGCTTCCTT GGAGCAGCCT CTCTGCGCAG AAAGCCATCA 420  
60 GGTGCTCAAT CATCTTCTCC TGGCCAAGGC TGTGACCATG CTAGTACTG GAATAGAGGT 480

	GGCCAGGCCC CCAGCGACTC TTCTTGGCCT GATGTTTGTC CTCACAGGCA TGCCACGTGG	540
	CCTGAGATGA TTCAGAACAA ATCATGCTAA CTTTGAATCC ATCCAGCCAC TTGCAATGA	600
5	TAATCAGAAG TCAGCTTGTT CACTGTTAGA AAGAACTAA CAAAAGAGAA CCCAGAGCAA	660
	TCTAGAATCT TTGAGTGCTT GGCTTTCCAA GGATACTGCG GAGACTCTGG CCAAGCTGAT	720
10	GAMCTTCTGA ARTGTCACTG GCACCATATG CAACAAGAAC CACCATTAC TGAGTAGCTA	780
	ATGGGTTTGG GGCCTGGGAC ATTCCATCTG AGGTCCCTTC TGAACATGTC ACTCCACAGC	840
	AGAGGACCGG TTGCAGCTTA CCCAGAACCA CTCCTCCAGG AGAGCTGGAT GTTTTGCCTG	900
15	CAACACCTTG AGCACTGACT GCTATTGTTT AAAAAAGCC TTTGCTGCAT TCGGAGGACT	960
	GCCCCGTGCC CTGAGGTGAC TTCCTAATA TGTGGTTTCA TTAGCGAATT TATTTTGT	1020
20	GCTGGGTGGA CATTTGTATT TTGTTAGTTT GCTGTTTAAG CTCAGTTTG CTGTGCTCTC	1080
	TGCAGCTACA AAACATCTTG GCATATTTAA GAKTGGCTTT TATAAATAGC TTTATTCTGA	1140
	TATTAATCAG ATTCCCAACT TTAAGGACTG GGGTACTTTA AAGAAATGCA	1200
25	AATAGCAATT GAAGAACCAC TGCTGCAGGT GGTAGCCCTG GCTAGACTGA ATTACACTAG	1260
	AAATCAGCCA GAAGGAAGCG TCCTTGGGAT CCCAGATCAC TCTTTTTTTT TTTTTTTT	1320
30	AAAGGGGAG CCCCTTGATG GCTCATCTCT CTGAATAACA GTTACGTCTT CATATCGATA	1380
	CCAGATGCCT TCTTCATCAT GCCACTGAAG CCACTCACCA CCTCAAGAA CATGCCAACC	1440
	TCTGTGAGAT TCACTTACCC ACAACAAGG AGGCACGTTT GGCACAAAGT GTTGTCTCTC	1500
35	AGGTCCAAGT GGAAGCTACA GAGTGCTTGA CCTCAACACA CTGGATTCCA GGTGGACTGG	1560
	ACCAAGAGCA GGCAGAGACA CGGGAAGTGA AAAACTCCAC AGGGTTTGGG GAATAGAAAT	1620
40	GAAAAGCCAC GTCATATAAC TCAAGAATAA ATGGTGTGTT GGAAATTTTA AAATTATCAT	1680
	CGAAGGTGGT GAAACTATTT CAGGCCCAA TGAAAGGAAA TCGCCAGTTG GGGATGAAAT	1740
	CACAGAGCCT GTGTTTATG ATATGGTTGG ATGTCCACTG ATGAAATTTT AAAGGAGTTT	1800
45	CATTTTAAAA AGTGGCATG ATTCTACATA TGAGAATTCT TTAGGCCAAG AACTGTCT	1860
	TGGCTCAGAG GTGTTGGGAA TTAAAGCAGA GAGAAGCCAT TCGTGATGCT TAGAACCAAG	1920
50	GATGGTCATG TACACAAAGA CCATCGAGAC GGCCATTCTT GTTTACAAA CACTTACCAA	1980
	GAAAGCACTT TGTAGGGGAA CTATTAGTAAG TTCTTCTCAT TTCAATTATG TTCTTCCAG	2040
	GAAACAGGAG AGACTGAATT AATAATTCTC TCTTCTCTCT TAAGCACTTT TAAAATAATA	2100
55	AAGTACATCT TGAAATTTGG GGGGGCATCT CTGATTAAAA AAAAGAAAAA GGCTGCTTGA	2160
	TGTATGTTAT GCAGAGACAC TCTGCTCTG GTGGCTGCAG AGCAATACCC AAGCCTCAT	2220
60	TGGAAGGCTC AACATTGGA ATTGCACTTT AATTGATTAA TCCTCAATTC ATGTGGCCTT	2280

ACGGGATGGT GGGTCTGGGA CCCCAATTCA TTCTTATCTG CCAAGAATT ATCTAGAAGC 2340  
 ACATCAAATA CCAGCACCCC ACCTGCACAA TGGGGGTGGA AAACCTTTGT ATCCCTAAGC 2400  
 5 ATATTATTTT ATAGTGTCTG CCATGCCATG TGGAAATACT TTATTTTAA CCTCAGGATT 2460  
 TAAATAAAGT AAACACTATG ACATTTAAAA AAAAAAAAAA AAAACTCGAG GGGGGCCCGG 2520  
 10 TACCCA 2526

15 (2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1111 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

25 CACTGCACCA GCTTTGTTAT CTGTAAAATG ATGATAATAC CAACACCTTC TTCTTGGGGT 60  
 ACTGAAGATG AGAGAACATG ATATGTGTAA AGTGCCTTCC ACAATACCCA GAACATAGCA 120  
 AACATGTAAT GAATGTAGTA ATAGTAATTA TTTTATTTTC TTTTGATTCA GTTGGGACTA 180  
 30 TGTTCAGCTG TAACAGAATA CCCAAAATAA CTGTTTAAA CAAATTAAAG TTTWGTGTG 240  
 AAGTTTGTG ACGAATTCAG ACAATCCAGG GCTTTTATAG ATGCACCAGG ATCAGCAGGT 300  
 35 ACAAAGGCAT CTTTCCTGAT TTCTGCCAGT CTCAATGCAT GGGTTGCAAT CCAGARTCCA 360  
 RGATGGCAGT TCCAGCCCTG GTTAGGCCCA TATTAGCACA CAGAAAGAAA GAGAAAGGGA 420  
 TGTGCCTCTT CACTTTAATC ATAGTCCCA CTAGATGCAC CCACTACTTC TGCTGATACT 480  
 40 CCATTAGCTA ATGCTTGCTT ACATGGTCAC ACTTAGTTTC CAGAGAGACA TGTCTGGACA 540  
 GTCATGTGCT CAATTAATAT CCAAGTGTCC AATTACTGAG AAAAAAGAA ACTAGCACCT 600  
 45 TTGCTTGCTT GCATTCCTCT TAGCATAAGC CACATCTTT TTATGAAGTT GTCCTCAGTT 660  
 ACTTGGATGC CTCAGTTGTC CTTTCAWTTA GAAAWGCYCC TKGGACAYCC TGAAWCTGAC 720  
 TTCTTTTGTG ATCAGCACCA TCACTACCAC TGCCTCTTC AAAGCCACCA CGTTCTGTCC 780  
 50 CCAGGATGGT TGCAACAACC AQCATAGGGA CTTTTTGCCT TCTACTTCCA CACAATAGNC 840  
 CAGAGTAAGC TTTTGAAAAT GTAGGTCAGA TCATGTCTCT CTCTTCCTCT TCAAAACCCT 900  
 55 CCCGATGGCT TTTCATATTA CTCAAAAGAA AACCTAAAC TTTGCTGTGA GATCTATGTG 960  
 ACCCGGCTTA TTCTTCCTCT TACTTTATCT CTGTATGCT CTTCTCACT CTACTCCAGC 1020  
 60 CATCCACCT CTTGCTGCT TGTCTATAC TCCTAAAAGA AGTTCACTCT TCCCTTATGA 1080

TATTTGCACT TAAATAGAA AAAAAAAAAA AAAAAAACT CGAGGGGGGC C

1131

5

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 941 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

15

GGCAGGAGTA GCATTTTCATT TAATCTGCAG GTATATTCTC CCAACAGTTT ATTGTCATGT 60  
GATGTCTCTCA GCCAAGATTG TRAGGCAGAG AGGAGCTGTC CCAACCTACT ATACCACCGA 120  
20 GGCTGGAGAG ATCATATTTT TGGTATTAAA CTGGAGTCTC TCCATCCTTC ACATTGTTGA 130  
TGTCTCTGT AGCAAACCGG AAAAGTCAGT GACAGAAGAT GCGGCTAGCG GTTTGAGCCA 240  
GAGAATGACA GCTCTGTTTT GGAGAAAAGG GCCGGATGGT GGCTCTAGAA AGCCCATCCT 300  
25 TCTGCTCTTC TTTTCTCTCC CCTTATATT GTGCTTTCAT TCATTCATTC ATTCATCAA 360  
CATTTGTTGA GCACCTATTA TGTGTCAAGC TCTGTGCTAG CCTCTGGAAA ACCTGCCCTC 420  
30 ATGTAGCTCA CTGTGGAGTA GGAGAAACAA TGACTACACT ATGATAAGCA CGGTTGTCA 480  
GGGTCTCACA GAGCAGTGGC CCTCATCCA GACCGATGAG GTCARAGAAG GCATCCAGGC 540  
GAGGATGGTG TCAGAGCTAA CTGAAGAATG AGAGGGAGCT GCACCASCAG GGTTTGAAC 600  
35 TGAAGGTGGC AGTGCCTGGA GTCTTGATT CAGCAGAGGG AGAGCAGTCT GTGAAAAGGC 660  
ACCAAGGGTG GGAGAGGGCA GAGCACATGG AGGAACTTCA GGTAGTTCTG GATGGCCTG 720  
40 GGGCAAAGCT AGAGAGGTAA GAAGAATCTA CAAATGTTCC TCGAGTTACA TGAACCTCCA 780  
TCCCAATAAA CCCATTGGAA ACGAAAAATT TAAGTCAGAA GTGCATTTAA GGCTGGTCCG 840  
AGTAGAATGA TTTTACAAC GAATTGATCA CAACCACTTA CAGATGTCTT TGTTCCTTCT 900  
45 CCACTCCCAC TGCTTCACCT GACTAGCCTT TAAAAAAAAA A 941

50

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 843 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

60



5  
 10  
 15  
 20  
 25  
 30

CNAGGGATAA CCCCAAGNT GGGAAATAAA CCTCRAATTA AAGGGGGAAC CAAAAAGCTG 50  
 GGAAGTTCCT CCCCGCGGTG GCGGCCNGNT CTAGGAACCTA GTGGAATCCC CCGGGGCTGC 120  
 ACGGAATTCG GCACGGAGTG GGAATGTTGT TTGTATGATA CTATTTCCAC AAWATCCATT 130  
 GAGACTTGGT KTGTGGCCTA GGACATGGTC AATTCTTTT AAATATTCGG TGAATTTCTT 240  
 TAGTGCAATAT TCTCCGATGG GGGCTGTGGG GACAGAGTTC TAAATATGCC CATTAGATTA 300  
 AATCTCTTCA TTCTGTGCT CACATCTTCT ATATCCTTAT TAATCTGTCA ATCTCTTCAA 360  
 GAGAGTGTT ATTAAATCT CTCACTGTAT GTGTCACTTT GCCCTTAAAA TTCTGATGAT 420  
 TTGCTTTATA AATGGTTATA ACCATTTTCC AGGAAGAACA TTAAAGAACT TTCCATTGGC 480  
 ATTATCCAGT TTCCCTCAA ATACTGGTTT TTTTATTTT GGCTNCTAAG CAGCTATGAA 540  
 TCCAGTTTCT CAGAAGCCTT TGTCTCAAGG CATTTGTTTC CAGATTACCT TGTTAGCATC 600  
 CACACTATGG GCTATTTTAG AAAACAAAA AAAGTATCAA AATCATATAG CTATGATTTT 660  
 CCTGTGCTTG AAGGAGCCTT AAAGCTCATC TAGTCCAGCC AGTATTTGTT CATCCAAAT 720  
 CTGCCAAGAA ATCTCTATTG TCAAGATATT CTTTACCATC TTTGGGACAT TCTCATATT 780  
 AGAAACAAAT CCTAAGAAGA AATTCTGCCA TAKACAACCC ATCCGTTCTT TAAAAAATA 840  
 AAA 843

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1018 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

45  
 50  
 55  
 60

CTGTAATTTT TAATTTTCAT ATACCGTGCT TTGATTCTAA TTTATTTTT TGAGTCTCT 60  
 GAAGGTTACA TATACAGAGT GCTTCAGGAA TGATCATTTT GTTATTATTC ATGCTTCTTA 120  
 ACAATGTTGT TTTAGTCCAA GAAGATAATT GCCAGAGAAA GAATACAGTG CAGGAAAGAA 180  
 GARGCTGGAG CCAGTGGTGA AGARGGATTG AGARGACAGA CATTGTGGGA ATGAAATCAT 240  
 GAATAATCGT GTTTTTGAAT TGTCCAAAAA CTCTACAAA CCATGAAATG TTGGACTTTA 300  
 AATCTAATTG TTGAAAAAT CCCCACTTC CTGTATCCC TTAGGTTGAG CATAATTCCA 360  
 CATCCGTGGA CTGATGCACT TCCCAAGAGG GGGCTCATT AACTCTTCGG AGGCAGCAGC 420  
 AGCAAGGGCA CCCCCTCCTT TCCCCCACA CCCCAATTCT CATGGCTCTT CTTTCTCTCA 480  
 TCTCATGCTT AGGTTAGAAA AGGCGACAAG GTAAGGAAGC CTTTGGGAAT AGGCTGAATC 540

TGGCTATCTA ATTTGGTGCC AAATACTTAA TGTGCTTGAA TTTAAAAACA GCAAACATGT 600  
 AGAAAGGTAA TTATAATTAT GAGGCCAGTT CTTTAAGCTA GCTTTTTTTC CCCTCTCAAA 660  
 5 CAGCATATTG GCTTGGATGT CAGCAGGAGA AAGTGTTTTT TGCAATACAC ATAATGCATA 720  
 TATGGTCCTG TTAGCAATCT ATAGAAAATA GATATTGCTC ATTAAGGTAA ATATTTTTGT 780  
 10 TGATGAATGA TCTGGAATGG TCTGGACTTG TTGTGTGAAC AGGAAATTGC TCTGTAGGCT 840  
 TTGACTTGTG AGGTAAAGAG TGAGGCTGCT AAGATTAATT AAAGTAAATA CTGTGACAAT 900  
 AGGATGTCAA AACCAAAAAC GTGTTTCTGA AACTCAAGGA ATTAATGACA CATAGGGAAG 960  
 15 TTTTGTCCAT ATTAAGCATA GAGTAGGAGA GCCAAGTCAA GAATAAAAAA AAAAAAAA 1013

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 661 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TTTAAGAAAT TAGTGAATCC CCGGNTGCAG GGAATTCGGC ACGAGGAGGA GCCCGTCAGC 60  
 TGGCAGGAGC GCAGGATGGC AGCTGYTCCC CCGGGTTGCA CCCCCCAGY TCTGCTGGAC 120  
 35 ATAAGYTGTT TAACAGAGAG CCTGGGAGCT GGGCAGCCTG TACCTGTGGA GTGCCGGCAC 180  
 CGCCTGGAGG TGGCTGGGCC AAGGAAGGGG CCTCTGAGCC CAGCATGGAT GCCTGCCTAT 240  
 GCCTGCCAGC GCCCTACGCC C CTCACACAC CACAACACTG GCCTMTCCGA GCTGCTGGAG 300  
 40 CATGGAGTGT GTGAGGAGGT GGAGAGAGTT CGGCGCTCAG AGAGGTACCA GACCATGAAG 360  
 GTGCGCAGGG CAGGCTCGG ACCTACCCCA GGAATGTCTT GCCCTGGGAA TGACAACACA 420  
 45 GTCCACACCA TGCACGGGGA GGCAACAGG GGCAGCTGAC CCAGCCCAGG GGTGAGANGA 480  
 GGTCTTGCCG AGGAAGTGGC AGCTAAGCTG ATACCTGATA TGCACWAGKC AGCCARGYGG 540  
 AGACAGGCCAA GGAAGAAGCT TGTTTTGAGG ACAGAATTTT CTAGATCACT CAGCACCATC 600  
 50 TGGCTTTTGG GCCTTTTGT TTTATTTTGT TTTGAGACG GGTCTCGCT CTGTGCCCCA 660  
 N 661

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

275

- (A) LENGTH: 533 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GGCACAGGGC TATTTGCCCC TCTCTCCACA TGACAGAACT GCTCTAAGTT TCTTTGCTGC 60  
 10 TCTTCTCAGC TGTCAGACGG CTTGCTGCTT GTTTCCACA CCACCATGTC TATTCTTTGC 120  
 TGTCCTTAC TCTGCCTGTT TTTTCCTTT TGTATTTCTT CTGGCTCTTG TCCCTTTTCC 180  
 15 CACGTGTCWC AGCTTTCCTT TATTGCCACT TTCAGTCAGA GCAGTCCTGT GCTTCTGGTG 240  
 CCGGCATACA ATACTTACTT GAGTTTCTTG GCTTTTCTTG ACTGTGCATC TCTTACTTCA 300  
 ACATAGGAAT AGCCTGTCAT AGAATTTCTC CAGTTCACGG GCTCAAGAGG GAGAGTCCCA 360  
 20 GAAAATTGAG ACTGTTTCC CTGTCTTGA TTGAATTCAT AAAGCAAAC CAGTGTGTGT 420  
 GTGAGGGTTT GCTGTGTCAT GCCTATAGGT TGTGTGGTG CAAACCTATA GAATCCAGCC 480  
 25 TGCGAAAGA AAGRAACCAG AGAATANCAG CATCAGAACA ATGCTTGACA TCATTTCTCA 540  
 ATCAAGCAGT CCA 553

30

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 869 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

40

45

50

55

60

GGCAGGAGCT GCCAACACTG AGGTCTTCGT GGCTTCTCAC ATCTAGATGT ATCCCTCTCA 60  
 AATCTATCCT CTATCCAGGC ACCAGATTGA GGTATCTAAA ATGTCAACTT TCCAGTTACT 120  
 CTTCTTTATA CTAGCCCAAT CAACTTACAA GATAAAGTCC AAGCCCCTTC ATATGACAAA 180  
 CCACACCCCTG CTTAACTCTC CAGGTTTGAA TCCTTCATCT CTTACTTTAA ACTTTAAAAC 240  
 CCAGCAGCAC GAAAGTGTCT CCTATGCATG TTGCCATATG CGTTCTCTCC ATCATGCATT 300  
 TGCCCTGAGCA AGATGTCTTG AGTTAACATC TTATTTCTTA AGACTCATTG TGGTGGTAGA 360  
 CAGCCTTTAA TAACGGATCC TTGGCCAGGC ACAGTGACTC ACACCTGTAA TCCCAGAACT 420  
 TTGAAAGGCC AAAGAAGGAA GAAAGCTTGA GGCCAGTAGT TTGAGACCAG CCTGGGAAAC 480  
 AGAGAGATAT CCCATCTGTA CCAAAAATTT AAAAAATAT TAGCAGGGAG TAGTGGCATG 540  
 CACAAGTGGT CCCAGCTCCA TGGGAGASTG AGGTACGAAC ATCACTTGAG CCCAGGAAGT 600

276

CAAGGCTGCA GTGAACCATG ATCAGAACAT TGCANTCCAG CTTGGGTAAC AGAGTGAGAC 660  
CTTAGGTCAG AAAAATGAAT AAATAAGCAT AAAATTTTAA AACTTAGCC AGGCATGGTG 720  
5 GCACACATCT GTGGTCCCTG CTACTIONAGGA GGCTGAGGTG AGAGGATCCT TGAGCCCAGG 780  
AGGTCAACAC TACAGTGAGC TATGATTGTG CCACTAACT CCAACCTGGG TGAAAAAGCA 840  
10 AAACCTGCC AAAAAAAAAA AAAAAAAT 869

## (2) INFORMATION FOR SEQ ID NO: 19:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 959 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GGCGAGCCGA GATCGTGCCA TTGCACTCCA GCCTGGGCAA CAAGAGTGAA ACTCTGTCTC 60  
25 AAAAAAAAAA AATTATAATA CTATATGCCA TAAATGACA TTTCATATTT AAAGAGTTTT 120  
TTAAACTCT TGTATTCACA TGCCATAATT TGAAACCCTA TTCACTGAA TGAGAATGGT 180  
30 ATCTGTGTG CTCATTTTTT CATTTTTATC CTTAACAATT TCCACCACAG CCAGTGCCATA 240  
TAATGGCAAT GACACCCAGG GATGGAATGA TAAGTTCCAT CRCMGCTCAG TCAAGACGCA 300  
GACTTGATGT GGGCCCAACA ACAGTCAATA ATGGAGTCTC CAAAATAAAG CTCTATAGGA 360  
35 AAGGTAAATA CCGCTGCAC AAGAAACCAC AGCATCTAGG TTCTAACCCC ATCTCTATGA 420  
AGAGCTTGCT GGGAGAGTTT TGACATTWAA CAATCTGTCT GATKGCCAAT TTTTCTCTC 480  
40 TATAAATGA TAATGTTKGA YTCAAAGATC CAAAGTCAAT TCATGGTCTA AACTTAATG 540  
ATTTTTTTAG GTTTTGKAC ATTTCACTGT AACTGTAGT AATTTATATC TTATTTTCCC 600  
ACTAATTTAG AAAAATATYT AATGATCCT TAATGGCAA TGGGTCTTAA GAATTTTGT 660  
45 TTAAATCCCT GTTACCCAAA AGAGCCCTTT TTTGTATCTC GCAGTAGTTA CAAGGATCTT 720  
TCTAAATCTT AAAAAAAAAA AAAAAAGAAA GAAAGAAAAG AAAAGAAAAA AAGTCAGCCG 780  
50 GCGTGGTGG CTCATGCTG TAATCCCAGC ACTTTGGGAC CAAGGTGGAC AGATCAGCAG 840  
GTCAGGAGAT GGAGACCATC CCGGCCAACA TGGAGAAACC CTGTCTCTAC TAAAAAAAAA 900  
55 AAAAATCGA GGGGGGCCCC GTACCCAATN CGCCGGCTAG TGTCGTAAA ACAATCAAA 959

## (2) INFORMATION FOR SEQ ID NO: 20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1446 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

5 CCGGGCAGGG CTGTGTGGCA CCGCCAGGGA CCGGGCCAC CTGAGTCACT TTATTGGGTT 60  
 10 CAGTCAACAC TTCTTGCTC CCGTTTCTT CTCTGTGGG ATGATCTCAG ATGCAGGGGC 120  
 TGGTTMTGGG GTTTTCCTGC TTGTGGAAG GCGTGGACAC TGCTGGGGG CTGGAAGCC 130  
 15 CCTCCCTTCC TGTCTTCTG TGGCCTGCAT CCCCTCATGG GTGCTGCCAT CCTTCCTGGA 240  
 GAGAGGGAGG TGAAGCTGG TGTGAGCCA GTGGGTTCCT GCCACTCAC CCAGGAGCTG 300  
 GCTGGCCAG GACCGGAGA GCGAGCACTG CTGCCCTCTT GGGCTGCTC CTTCGGCAGT 360  
 20 TAGGGGTGGA CCGAGCCTCG CTTTCCCCAC TGTCTGGAG GGAAGGGGAA GGAGGGGGTC 420  
 TTCAGGCTGG AGCCAGGCTG GCGGTGCTGG GTGGAGAGAT GAGATTAGG GGGTGCCTCA 480  
 25 TGGGGTGGCC AGGCCTGGG TGAATRAGA AAGGCCAGA ACGTGCAGGT CTGCGGAGGG 540  
 GAAGTGTCTT GAGTGAAGGA GGGACCCCC ATCTGGGGG ATGCTGGGAG TGAGTGAGTG 600  
 AGATGGCTGA GTGAGGTTA TGGGAGCCT GAGGTTTAT GGGCTGTGT ATCCCCCTCT 660  
 30 CCGGGCCCCA GCCTGCTCC CTCTGCCC GCTGGCCAC AGGTCTCCT CTGGTCCCTG 720  
 TCCCTCTGGT GGTGGGGAT GGAGCGCAG CAGGGGTGT AATGGGGTG GGTCTGTCT 780  
 35 TCTACAGGCC ACCCCGAGGT CTTCACTGGT TGCCTGGGA GCGGACGGG GCTCCTGAGG 840  
 GGTACAGGTT GGTGGGGCC TCCGTGAGG TCTGGGTCA GGCTTTGGCT CTGCTGCCTC 900  
 TCAGTCACCA AGTCACCTCC CTCTGAAAAT CCAGTCCCTT CTTTGGATGT CTTGTGAGT 960  
 40 CACTCTGGGC CTGGCTGTG TCCCTCCTCA GCTTCTGTT CCTGGGACAA GGTCAAGCC 1020  
 AGGATGGGCC CAGGCTGGG ATCCCCACC CCAGGACCCC CAGGCCCTT CCGCTGCTGC 1080  
 45 TTTGGGGGG GCAGGGAGA AATGGAATCC TTTGGGTCC CCGAGGTGG GTCCCTCCC 1140  
 AGCCCTGCAT CTTCCGTGCC STAGACCTGC TCCCLAGAG AGGGGCTTG ACCCACAGGA 1200  
 CGTGTGGTGG CGCTGGCAC TCAGGGACCC CCAGCTGCC CAGCCCTGGT CTCTGGCGCA 1260  
 50 TCTCTTCCCT CTTGTCCGA AGATCTGCC CTCTAGTCCC TTTGAGGGG TTCCCATCAT 1320  
 CCTCCCTGA TATTGTATG AAAATATTAT GCACACTGTT CATGCTTCTA CTAATCAATA 1380  
 55 AACGCTTTAT TTAAAGCAA AAAAAAAAAA AAAAAACTCG AGGGGGGGCC CTTACCCAAT 1440  
 TCGCCA 1446

## (2) INFORMATION FOR SEQ ID NO: 21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1471 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

5	CAAAAATAA TAATGATAAT TTAATAATAA TAAGTAACTA ATAAAAAGAT TTTATATCCC	60
15	AGTCTTATGA TGTGGTTGG CAAGGCTAGA TAAAAAGATG TTAGAATGAA AGAACATATT	120
	TTTACTGATA TGTAPATGAA GGATTCTACA ATAGTCATAT ATTTTATAT GAATGAATGT	180
	TGGGTTGGGC TGGAGAGGTA TGTGTGTGTA AATATAAAGG TCTCACATTC AGAGTATACC	240
20	TCTGAAATAA TGGAACTCAT GTCTACAATT CAACATGCAT CTGTATAGTT ACATCTCATG	300
	TAAATATACA CAGACATATT TTGCAGCCAG TAATTGACAG TTAATGTCCA AAACAGGTGA	360
25	TTGATAGGTA ACAGAAATTA GATAACCACC AATTTTGGCC AAGAGAAAAG CTAGAAGGAC	420
	TAAAGCAGT TGAATGTATG GTACTGACAT TGTCTAAGC AGTCTGATAA CCAGTTTATT	480
	GAAACGTGTG CATTACAGA GAATTTAATT TTAACCCAT AATTTCTCCT ATCCATTAAA	540
30	ATATTATAAT TGTAGTAGT ATGAAACCAA CAGGAAATGT TTTTAACTA TTTAGTGAGG	600
	TGATTCATTT GTTTCATGGG CAAACACTAT CCAGGAAAAG CCTTGCTTGC CTGTTTCCCA	660
35	AAGAGCTCTA AGAAATAGAA TCAAGTGTA AATGGTTCAG ACCATTTCAGG ATTTCTTGTC	720
	ACTCTTCTCA ACCCCGATCT TCCTGTATT ACTGATGTTT GAAACCCTGT CATTAGCCCC	780
	GGCCTGGTTA AAGCCCTCA GAGTCACCTC TCATTCATAG CAATAGAATT CAACCCCAAG	840
40	TGGTTGATGG TGTCCCCAGC ACAGCCGAGA GACCTGATCT CTGGATTGAG TGCTTTTAGC	900
	TCTTCGAGTT TACCCTAAGA TACCTTCGGG CAATATTTT AACCAACCCA AAAGCTCTTC	960
45	AGGTCATTTT TGAAGAGGAC AAGGTGAATC TTGGCTTGGG ACACCATTTT TGGGCTCTTG	1020
	CTACTGAATG AATCAGAAAG GAATTTTTC TGAAGAGCAT TAGAAAGTAA AGGAGATGTT	1080
	AAAATAAGTT CTTGAAGTAT GTTTTATATT TATCTAAAAC ACTGATTTTA AAAGTTTACA	1140
50	TTCAAATGTG TATTCAAAAG AAGTACTGAT TTGTAATTAT TATAGTTTGT GTGTATCATC	1200
	CCCTTTTAAC CGTGCCTAAC AACTGTACTT AAATTTTGTT TTCCTAGTGT AACAAATGTT	1260
55	TCCCTAAGA TTTTCTAGAG CCAAATAATG GGAGTGAAAA ATTCCTTAAG TGTATATATA	1320
	GAAATATAT TAGAAAATCA GCTTTGGATT ATACGATTTT TAAAATATAC TAATACAGAA	1380
	TCCTCAGTAA TATGTTTGA ATTGGATTTT TTCTCAGAAC TGTTACATTA TAAATAATAC	1440
60	ATCAACCAGA AAAAAAAAAA AAAAAAATTN C	1471

## 5 (2) INFORMATION FOR SEQ ID NO: 22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1402 base pairs  
(B) TYPE: nucleic acid  
10 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

15 AGGGACGTCT TGCCTGAGGA GATGCCCATT TCTGTCTTGG RTTACCCCTCA CTGCGTGSTG 60  
CATGAGCTGC CAGAGCTGAC GCGGAGAGT TTGGAAGCAG GTGACAGTAA CCAATTTTGC 120  
TGGAGGAACC TCTTTTCTTG TATCAATCTG CTTCCGATCT TGAACAAGCT GACAAAGTGG 180  
20 AAGCATTCAA GGACAATGAT GCTGGTGGTG TTCAAGTCAG CCCCCATCTT GAAGCGGGCC 240  
CTAAAGGTGA AACAGCCAT GATGCAGCTC TATGTGCTGA AGCTGCTCAA GGTACAGACC 300  
25 AAATACTTGG GCGGGCAGTG GCGAAAGAGC AACATGAAGA CCATGTCTGC CATCTACCAG 360  
AAGGTGCGGC ATCGGCTGAA CGACGACTGG GCATACGGCA ATGATCTTGA TGCCCGGCCT 420  
TGGGACTTCC AGGAGAGGA GTGTGCCCTT CGTGCCAACA TTGAACGCTT CAACGCCCCG 480  
30 CGCTATGACC GGGCCACAG CAACCTGAC TTCTGCCCAG TGGACAAC TGCTCAGAGT 540  
GTCCTGGGCC AACGGGTGGA CCTCCCTGAG GACTTTCAGA TGAACATGA CCTCTGGTTA 600  
35 GAAAGGGAGG TCTTCTCCAA GCCATTTC TGGGAAGAGC TGCTGCAGTG AGGCTGTTGG 660  
TTAGGGGACT GAAATGGAGA GAAAGATGA TCTGAAGGTA CCTGTGGGAC TGTCTAGTT 720  
CATTGCTGCA GTGCTCCCAT CCCCCACCAG GTGGCAGCAC AGCCCCACTG TGTCTTCCGC 780  
40 AGTCTGTCTT GGGCTTGGGT GAGCCCAGCT TGACCTCCCC TTGTTCCCA GGGTCTGCT 840  
CCGAAGCAGT CATCTGTGCC TGAGATCCAT TCTTCCTTTA MTCCCCCAM CCTCCTCTCT 900  
45 TGGATATGGT TGGTTTGGC TCATTTTACA ATCAGCCCAA GGTTGGGAAA GCTGGAATGG 960  
GATGGGAACC CCTCCGCCGT GCATCTRAAT TTCAGGGGTC ATGCTGATGC CTCTCGAGAC 1020  
ATACAAATCC TTGCCTTTGT CAGCTTGCAA AGGAGGAGAG TTAGGATTA GGGCCAGGGC 1080  
50 CAGAAAGTCG GTATCTTGGT TGTGCTCTGG GGTGGGGGTG GGGTGTCTCT CATGTTATTC 1140  
CAGCCTCCTG CTACATTATA TCCAGAAGTA ATGCGGGAGG CTCCTTCAGC TGCCTCAGCA 1200  
55 CTTTGATTTT GGACAGGGAC AAGGTAGGAA GAGAAGCTTC CCTTAACCAG AGGGGCCATT 1260  
TTCTCTTTTG GCTTCGAGG GCCTGTAAAT ATCTATATAT AATTCTGTGT GTATTCTGTG 1320  
60 TCATGTGGG GTTTTAAATG TGATTGTGTA TTCTGTTTAC ATTAAAAAGA AGCAAAAATA 1380

ATAAAAAAAA AAAAAAAAAA CT

1402

5

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 1047 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GGCACAGGGG	ACTACAGGCA	CCCACGACCA	TACCCAGCTA	ATTTTGTAT	TTTTTGTAG	60
AGATGGGGTT	TCACGATGTC	GCCCAGGCTG	GTCTTGAAC	CCTGGGCTTG	AGCGATCTTC	120
CCATCTTTCC	ATCTTGGCCT	CCTAAAGTGC	TGGGACTGCA	GGCATGAGCC	ACCATGCCCA	180
GCCAAGATTC	TTATTGATTA	CCATGTTGCT	TCAAGAAGCC	AAGCCAGTTT	CCAATATTCC	240
CCATTTGCTG	GAGTCTTGGT	ACTTTGGGTA	GAAGCAACTG	GTAAATTGTT	AATTGGAACA	300
NTTGGTGGTG	TAGATAACCA	CGTATGGCCA	AACCTAGAGC	ATCTAGGCTC	ACAATTACTA	360
TCCTGACTTG	ATAACAAGTG	TTCTGATATT	AACCTGAAAA	TGGGAATAAT	GCCAAATCTG	420
TGTAAC TTAA	CATCTATATA	CACAGTGGGG	AGAAGTGAAG	TTATTAAACC	TGGAATCTCT	480
GTGATCAAGG	CTAACAGTAG	TTATCTAAGA	AGCAAAGGAC	CTACAATTCT	TAGACTTGGA	540
GTCATATTCT	TTAAGGACGT	GTCTGAAAC	TATATCAAGC	ATCTGGTTTC	CACGTATTTT	600
TCCCTCAGAA	ATTATGAAGT	ACAAGTAAAA	ATGAAGGTAC	AGGGTAAGAC	ACATGCTGCT	660
TTCTTGCTCT	TGAGTGGAGA	CAGTTTTC	CCCATCTTAA	CCCCTTWACA	CAAAACAATT	720
TGTGTTTTAT	AGCAAATAAG	TGACTCAACA	TAATTTCAAT	ATGATGTTTA	TCCACCAGTA	780
CTTTCCTTTC	AGCTTCTAGT	CCCATAARTG	GTTTGTGAAG	TCATCGGTTA	CATTAGCCAA	840
GATAGGCCTA	GACTTGAAGT	CTAGAATGTT	TTTCCCACTA	TATGCCAAAG	TAGAATGTGG	900
GTATCTCAGG	GTCATTTTTC	TTGTTCAATT	TCCCACCTGT	ACAGTTGTTA	TGATTCACCT	960
TCCTTATGTG	TCTAATAAAT	CTTGTTCCAT	GAAATGATCA	AAAAAAAAAA	AAAAAAACT	1020
CGAGGGGGGG	CCCCGTACCC	AAATCGC				1047

55

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

60

(A) LENGTH: 990 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double



## (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

5 TTGGAAAGGG TCTAGCTCTT TCTCATTCAC CAACTATATT AGAAGCACTT GAGGGAAATT 60  
 TACCACTCCA AATCCAAAGC AATGAACAGT CTTTCTCGGA TGATTTTATT GCCTGTGTCC 120  
 10 CAGGATCAAG TGGTGAAGG CTTGCAAGGT GGCTTCAGCC AGATTCATAT GCGGATCCTC 180  
 AGAAAACATC TTTGATCCTG GAATAAGGAT GATATTCGTT GTGGTTGGCC TACCACCATA 240  
 ACTGTTCAAA CAAAAGACCA GTATGGGGAT GTGGTACATG TTCCCAATAT GAAGGTAATT 300  
 15 ATAACCTGGAT TAAATTAGCA GACATCTATA TACTGGCTGC AATGACTGAT AAAATTTTAG 360  
 AAATGCCAAG TGCTGAGGT CCATTGTTC TACCTCTTT ATATAAAGG TGATGCTGAA 420  
 20 AGTTTGTTTA AATGACTGT TTATATTAAT TAGTCCCAA GTGTCCAAGT TACACCTGTT 480  
 TTTTGTGTA GTTTGTTCTT TACATTTGC TACCTGTTAC GGGGACTCAA AGGAGGGATA 540  
 AGAAAGTATC CATCTAAAGA GTGCTAGACA CATACAGTGA AGCCCTCAA TATGTATTGA 600  
 25 TTGAATAAAT GCATGAAAGA ATACATTTTT AAATTTTGTG TATAGTTTTC AAAGACTCAA 660  
 GTACGTTCTG TGTGTTGTTAT TACTGAAACC ACATTTTAAA AATAACACTC ATTAAGTTAG 720  
 30 AAATATATGA GTTTAGATTG TAAAAGAATG AGGAATTGAA ATAGTTGTAT ACCATATTGA 780  
 TGAATATAGA GTTTTTAGGA TACCTCTTAC CTGAAATATT AATAATAATG TTTNCAGACC 840  
 ATATTATACA TAATTATTG TGATTTAATC TGTTAATATG AATATCTCAT TTAACCTTT 900  
 35 TATTTCTGAA AAAATTATAT TGAATAAAAT TTTATATAGG CAGTCCCCAG CCCTTTCCTC 960  
 CTTCAAAGTT GTCTTATAGA GTGATTGGTT 990

## (2) INFORMATION FOR SEQ ID NO: 25:

## (i) SEQUENCE CHARACTERISTICS:

45 (A) LENGTH: 1208 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

50 TAATCGCTAC TATAGGGAAA GCTGGTCGCT GCAGGTACCG GTCCGGAATT CCGGGTCGAC 60  
 CCACGCGTCC GAGCGAAATG GCGCTCCGG CCCCCGGCCC GGCTCCGGC GGCTCCGGG 120  
 55 AGGTAGACGA CCTGTTGAC GTAAAGAACG CTTTCTACAT CGGCAGCTAC CAGCAGTGCA 180  
 TAAACGAGGC GCASGGGTGA AGCTRTCAAG CCCAGAGAGA GACGTGGAGA GGGACGTCTT 240  
 60 CCTGTATAGA GCTTACCTGG CGCAGAGGAA GTTCGGTGTG GTCTGGATG AGATCAAGCC 300

CTCTCGGGC CCTGAGCTCC AGGCCGTGGC CATGTTTGCT GACTACCTCG CCCACGAGAG 360  
 TCGGAGGGAC AGCATCGTGG CCGAGCTGGA CCGAGAGATG AGCAGGAGCK TGGACGTGAC 420  
 5 CAACACCACC TTCCTGCTCA TGGCCGCTC CATCTATCTC CACGACCAGA ACCCGGATGC 480  
 CGCCCTGGCT GCGCTGCACC AGGGGGACAG CCTGGAGTGC ACAGCCATGA CAGTGCAGAT 540  
 10 CCTGCTGAAG CTGGACCGCC TGGACCTCGC CCGGAAGGAG CTGAAGAGAA TGCAGGACCT 600  
 GGACGAGGAT GCCACCCCTCA CCCAGCTCGC CACTGCCTGG GTCAGCCTGG CCACCGCTGG 660  
 TGAGAAGCTG CAGGATGCCT ACTACATCTT CCAGGAGATG GCTGACAAGT GCTCGCCAC 720  
 15 CCTGCTGCTG CTCATGGGC AGGCGGCTG CCACATGGCC CAGGCGCGCT GCGAGGCGGC 780  
 TGAGGGCTG CTGCAGGAGG CGCTAGACAA GGATAGTGGC TACCCGAGA CGCTGGTCAA 840  
 20 CCTCATCGTC CTGTCCGAGC ACCTKGGCAA GCGCCCTGAG GTGACAAACC GATACCTGTC 900  
 CCAGCTGAAG GATGCCACA GGTCCCATCC CTTTCATCAAG GAGTACCAGG CCAAGGAGAA 960  
 CGACTTTGAC AGGCTGGTGC TACAGTACGC TCCAGCGCT GAGGCTGGCC CAGAGCTGTC 1020  
 25 AGGACCATGA AGCCAGGACA GAGGCCAGGA GCGAGCCCTG CAGCCCTCCC CACCCGGCAT 1080  
 CCACCTGCAT CCCTCTGGGG CAGGAGCCCA CCCCCAGCAC CCCCATCTGT TAATAAATAT 1140  
 30 CTCAACTCCA RGTGTTCCA CCTGAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1200  
 AAAAAAA 1208

## (2) INFORMATION FOR SEQ ID NO: 26:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1922 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GTGCTGGCT ACTGAGCAGC GGCATGGAGG ACTCTGAAGC ACTGGGCTTC GAACACATGG 60  
 GCCTCGATCC CCGGCTCCTT CAGGCTGTCA CCGATCTGGG CTGGTCGCGA CCTACGCTGA 120  
 50 TCCAGGAGAA GGCATCCCA CTGGCCCTAG AAGGGAAGGA CCTCCTGGCT CGGGCCCGCA 180  
 CGGGCTCCGG GAAGACGGCC GCTTATGCTA TTCCGATGCT GCAGCTGTTG CTCCATAGGA 240  
 55 AGGCGACAGG TCCGCTGGTA GAACAGGAG TGAGAGGCCT TGTCTTGT CTACCAAGG 300  
 AGCTGGCAGG GCAAGCACAG TCCATGATTC AGCAGCTGGC TACCTACTGT GCTCGGGATG 360  
 TCCGAGTGGC CAATGTCTCA GCTGCTGAAG ACTCAGTCTC TCAGAGAGCT GTGCTGATGG 420  
 60

	AGAAGCCAGA TGTGGTAGTA GGGACCCCAT CTCCCATATT AAGCCACTTG CAGCAAGACA	480
	GCCTGAAACT TCGTGACTCC CTGGAGCTTT TGGTGGTGA CGAAGCTGAC CTTCTTTT	540
5	CCTTTGGCTT TGAAGAAGAG CTCAGAGTC TCCTCTGTCA CTGCCCCGG ATTTACCAGG	600
	CTTTTCTCAT GTCAGCTACT TTTAACGAGG ACGTACAAGC ACTCAAGGAG CTGATATTAC	660
10	ATAACCCGGT TACCCTTAAG TTACAGGAGT CCCAGCTGCC TGGGCCAGAC CAGTTACAGC	720
	AGTTTCAGGT GGTCTGTGAG ACTGAGGAAG ACAATTCTCT CCTGCTGTAT GGCCTGCTCA	780
	AGCTGTCAAT GATTGGGGC AAGTCTCTGC TCTTTGTCAA CACTCTAGAA CGGAGTTACC	840
15	GGCTACGCCT GTTCTTGGAA CAGTTCAGCA TCCCCACCTG TGTGCTCAAT GGAGAGCTTC	900
	CACTGGCCTC CAGGTGCCAC ATCATCTCAC AGTTCAACCA AGGCTTCTAC GACTGTGTCA	960
20	TAGCAACTGA TGCTGAAGTC CTGGGGGCC CAGTCAAGG CAAGCGTCGG GGCCGAGGGC	1020
	CMAAGGGGA CAAGGCTCT GATCCGGAAG CAGGTGTGGC CCGGGGCATA GACTTCCACC	1080
	ATGTGTCTGC TGTGCTCAAC TTTGATCTTC CCCCACCCC TGAGGCCTAC ATCCATCGAG	1140
25	CTGGCAGGAC AGCAGCGGCT AACAACCCAG GCATAGTCTT AACCTTTGTG CTTCCCACGG	1200
	AGCAGTTCCA CTTAGGCAAG ATTGAGGAGC TTCTCAGTGG AGAGAACAGG GGCCCCATTC	1260
30	TGCTCCCTTA CCAGTTCGG ATGGAGGAGA TCGAGGGCTT CCGCTATCCC TGCAGGGATG	1320
	CCATGGCTC AGTGACTAAG CAGGCCATTC GGGAGGCAAG ATTGAAGGAG ATCAAGGAAG	1380
	AGCTTCTGCA TTCTGAGAAG CTTAAGACAT ACTTTGAAGA CAACCTAGG GACCTCCAGC	1440
35	TGCTGGGCA TGACCTACCT TTGCACCCCG CAGTGGTGAA GCCCCACCTG GGCCATGTC	1500
	CTGACTACCT GGTTCCTCT GCTCTCCGTG GCCTGGTRCG CCTCACAAG AAGCGGAAGA	1560
40	AGCTGTCTTC CTCTGTAGG AAGGCCAAGA GAGCAAAGTC CCAGAACCCA CTGGCAGCT	1620
	TCAAGCACAA AGGAAAGAA TTCAGACCCA CAGCCAAGCC CTCCTGAGGT TGTGGGCCT	1680
	CTCTGGAGCT GAGCACATTG TGGAGCACAG GCTTACACCC TTCTGGACA GGCCAGGCTC	1740
45	TGGTGTAC TGCACAGCT GAACAGACAG TTCTGGGGC GGCAGTGTG GGCCCTTTAG	1800
	CTCCTTGGCA CTTCCAAGCT GGCATCTTG CCCTTGACAA CAGAATAAAA ATTTAGCTG	1860
50	CCCCAAAAA AAAAAAAAAA AAAAAAATC GAGGGGGGG CCGTACCCAA TTCGCCCTAT	1920
	AA	1922

55

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1951 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

5	TCGTCCCCAG AGCGGGCTGA GCCCCAGGCG SAGGGTGGCG GGGGAGCCTG GGGGAGCCGC	60
	CGCCACCTCC ACGGGCCTCT CTGAGCTCGG ACACCAGCGC CCTGTCTAT GACTCTGTCA	120
10	AGTACACGCT GGTGGTAGAT GAGCATGCAC AGCTGGAGCT GGTGAGCCTG CCCCCTGCTT	180
	CGGAGACTAC AGTGACGAGA GTGACTCTGC CACCGTCTAT GACAACTGTG CCTCCGTCTC	240
	CTCGCCCTAT GAGTGGGCCA TCGGAGAGGA ATATGAGGAG CCCCCGGGGC CCCAGCCCCC	300
15	TGCCCTGCCTC TCCGAGGAAC TCCACGCTG ATGAACCCGA CGTCCATTTT TCCAAGAAAT	360
	TCCTGAACGT YTTCTAGAGT GCGCGCTCCC GCTCCTCCAG TGCTGAGTCC TTCGGGCTGT	420
20	TCTCCTGCAT CATCAACGGG GAGGAGCAGG AGCAGACCCA CCGGGCCATA TTCAGGTTTG	480
	TGCCTCGACA CGAAGACGAA CTTGAGCTGG AAGTGGATGA CCCTCTGCTA GTGGAGCTCC	540
	AGGTGAAGA CTACTGGTAC GAGGCCTACA ACATGCGCAC TGGTGCCCCG GGTGTCTTTC	600
25	CTGCCTATTA CGCCATCGAG GTCACCAAGG AGCCCGAGCA CATGGCAGCC CTGGCCAAAA	660
	ACAGTGAAGT GGTGGACCAG TTCCGGGTGA AGTTCTGGG CTCAGTCCAG GTTCCCTATC	720
30	ACAAGGGCAA TGACGTCTC TGTGCTGCTA TGCAAAAGAT TGCCACCACC CGCCGGCTCA	780
	CCGTGCACTT TAACCGCCCC TCCAGCTGTG TCCTGGAGAT CAGCGTGGG GGTGTGAAGA	840
	TAGGGGTCAA GCGCGATGAC TCCAGGAGG CCAAGGGGAA TAAATGTAGC CACTTTTTC	900
35	AGTTAAAAAA CATCTCTTTC TGCGGATATC ATCCAAAGAA CAACAAGTAC TTTGGGTTC	960
	TCACCAAGCA CCCCCCGGAC CACCGGTTTG CCTGCCACGT CTTTGTGTCT GAAGACTCCA	1020
40	CCAAAGCCCT GGCAGAGTCC GTGGGGAGAG CATTCCAGCA GTTCTACAAG CAGTTTGTGG	1080
	AGTACACCTG CCCCACAGAA GATATCTACC TGGAGTAGCT GTGCAGCCCC GCCCTCTGCG	1140
45	TCCCCCAGCC CTCAGGCCAG TGCCAGGACA GCTGGCTGCT GACAGGATGT GGCAGTGCTT	1200
	GAGGAGGGGC ACCTGCCACC GCCAGAGGAC AAGGAAGTGG GCGGCTGGCC CAGGGTAGGG	1260
	GAGGGTGGGG CAATGGGGAG AGGCAAATGC AGTTTATTGT AATATATGGG ATTAGATTCA	1320
50	TCTATGGAGG GCAGAGTGGG CTGCCTGGGG ATTGGGAGGG ACAGGGCTTG GGGAGCAGGT	1380
	CTCTGGCAGA GAAGGATGTC CGTTCCAGGA GCACACGGCC CTGCCCCATC CTGGGCCCTTA	1440
	CCTCCCTGTC CAGGGCTCGG GCGCTGTGGC TCCTGCCTTG ATGAAGCCCC TGTCTGCTT	1500
55	TGATGAAGCC TGTGCCACCT GCAAGTGCCC GCGCTGCCCC TGCCCCAACC CCCACCGAAG	1560
	AGCCCTGAGC TCAGGCTGAG CCCAGCCACC TCCCAAGGAC TTTCCAGTGA GGAAATGGCA	1620
60	ACACGTGGAG GTGAAGTCCC TGTCTCAGC TCCGTCTCTT GCGGGGCTTC TGGGTGGCTC	1680

CTGCCACTGA CTCACCGGCC ATGCTGGCCT GTGGCAGGCC TAGGACCTCA GCGGGGAGG 1740  
 AGGAGCTGCC GCAAGGCCCT GTCCAGCAG AAGAGGGAGG CTTCCTGACT GACACAGGCC 1800  
 5 AGCCCCATCT TGCTCTGTC ACCCTGGCCC CAACTATTAA AGTGCATTT CCTGTCAAAA 1860  
 AAAAAAAAAA AAAATCGGGG GGGGCCCGGA ANCCAAATTC CCCCRAAAG GGGGGTTATA 1920  
 10 AAAATTCCCN GGCNGTGTTC TAAAAATTC G 1951

15 (2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3989 base pairs.

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

25 GGCACAGGCC GCAGGGNACC TATGGGGCCA TATAGTTGT AATGAACTG TAGTCTCAGT 60  
 TGGAAAGCTA GACATGAAAT GGGTCAGTGA GCAAGGCTCT ATTCTAGTC TCCAGCCATG 120  
 30 CCTGTGGAAC CTGARCCRC TCTCAGCACA TTGGACCCAG GCAGATGAA AAAATTCACA 180  
 GAACTATGAT TTGCACTCAA GGGTTGTAG ATTTCTCTCT TCATTCTAAT TTCAGTGTCT 240  
 AAAATCTCTG CATCCRTGAA CGAGCTGGGC ATTTGATGAG ACAGGGCTGA ATACTGCAGT 300  
 35 TTTCTCTCTA GAAATCATCT GGGCATTTT CTTTGAAGT ATGGGAACAA TAAGGCATAA 360  
 CTGTTTGACAC AAAGTTGGA TAATGATTT TGGGATAACG ATCTACCAGA ATGGGGATAT 420  
 TTCACCCCTG GTTCTGAGAT GCAACCAAA GAATATCATG ACCAGCTTTC AGGCCTCTG 480  
 40 AAGTATATCT CTCACATTGT CTGTTCTCA TGCTGAGGAG CCTGAGATCC CTGTGTGGG 540  
 ATTAGACAGT GCACTGTTAT GGGTGTAGT GAATTGGCTT ATTTGTCTG TCCCTGTCTG 600  
 45 AATGTATTGC AGGAATYAAA AAGGACCAAG AAGAGGAAGA AGACCAAGGC CCACCATGCC 660  
 CCAGGCTCAG CAGGGAGCTG CTGGAGGTAG TAGAGCCTGA AGTCTTGCAG GACTCACTGG 720  
 ATAGATGTTA TTCAACTCCT TCCAGTTGTC TTGAACAGCC TGAATCCTGC CAGCCCTATG 780  
 50 GAAGTTCTTT TTATGCATTG GAGGAAAAAC ATGTTGGCTT TTCTCTTGAC GTGGGAGAAA 840  
 TTGAAAAGAA GGGGAAGGGG AAGAAAAGAA GGGGAAGAAG ATCAAAGAG GAAAGAAGAA 900  
 55 GGGGAAGAAA AGAAGGGGAA GAAGATCAA ACCCACCATG CCCCAGGCTC AGCAGGGAGC 960  
 TGCTGGATGA GAAAGRGCTT GAATCTTGC AGGACTCACT GGATAGATGT TATTCAACTC 1020  
 60 CTTCACTTGT GTTGAAGTGT GTGACTCATG CCAGCCCTAC AGAAGTGCCT TTTATGTATT 1080

	GGAGCAACAG CATGTTGGCT TGGCTCTTGA CATGGATGAA ATTGAAAAGT ACCAAGAAGT	1140
	GGAAGAAGAC CAAGACCCAT CATGCCCCAG GCTCAGCAGG GAGCTGCTGG ATGAGAAAGA	1200
5	GCCTGAAGTC TTGCAGGACT CACTGGATAG ATGTTATTGG ACTCCTTCAG GTTATCTTGA	1260
	ACTGCCTGAC TTAGGCCAGC CCTACAGCAG TGCKGTTTAC TCATTGGAGG AMCAKTACCT	1320
10	TGGCTTKKCT CTTGACGTGG ASAAATTGAA AAGAAGGGGA AGGGGAARAA AAGAAGGGGA	1380
	AGAAGATCAA AGAAGGAAAG AAGAAGGGGA AGAAAAGAAG GGAAGAAGA TCAAAACCCA	1440
	CCATGCCCCA GGCTCAGCAG GGAGCTGCTG GATGAGAAAG GGCTGAAGT CTTGCAGGAC	1500
15	TCACTGGATA GATGTTATT CACTCCTTCA GGTGTCTTTG AACTGACTGA CTCATGCCAG	1560
	CCCTACAGAA GTGCTTTTTA YRTATTGGAG CAACAGYGTG TTGGCTTGGC TGTTGACATG	1620
20	GATGAAATTG AAAAGTACCA AGAAGTGGAA GAAGACCAAG ACCCATCATG CCCCAGGCTC	1680
	AGCAGGGAGC TGCTGGATGA GAAAGAGCCT GAAGTCTTGC AGGACTCACT GGATAGATGT	1740
	TATTGACTC CTTCAGGTTA TCTTGAAC TGCTGACTTAG GCCAGCCCTA CAGCAGTGCT	1800
25	GTTTACTCAT TGGAGGAACA GTACCTTGGC TTGGCTCTTG ACCTGGACAG AATTAAAAAG	1860
	GACCAAGAAG AGGAAGAAGA CCAAGGCCCA CCATGCCCCA GGCTCAGCAG GGAGCTGCTG	1920
30	GAGGTAGTAG AGCCTGAAGT CTTGCAGGAC TCACTGGATA GATGTTATT CACTCCTTCC	1980
	AGTTGTCTTG AACAGCCTGA CTCCTGCCAG CCTATGGAA GTTCTTTTTA TGCATTGGAG	2040
	GAAAAACATG TTGGCTTTTC TCTTGACGTG GGAGAAATTG AAAAGAAGGG GAAGGGGAAG	2100
35	AAAAGAAGGG GAAGAAGATC AAMGAAGRAA AGAAGAAGGG GAAGAAAAGA AGGGGAAGAA	2160
	GATCAAAACC CACCATGCCC CAGGCTCAAC GGCCTGCTGA TGGAAAGTGA AGAGCCTGAA	2220
40	GTCTTACAGG ACTCACTGGA TAGATGTTAT TCGACTCCGT CAATGTACTT TGAAGTACCT	2280
	GACTCATTC AGCACTACAG AAGTGTGTTT TACTCATTTG AGGAACAGCA CATCAGCTTC	2340
	GCCCTTTACG TGGACAATAG GTTTTTTACT TTGACGGTGA CAAGTCTCCA CCTGGTGTTC	2400
45	CAGATGGGAG TCATATTCCC ACAATAAGCA GCCCTTASTA AKCCGAGAGA TGTCATTCCT	2460
	GCAGGCAGGA CCTATAGGCA MGTGAAGATT TGAATGAAAG TACAGTTCCA TTTGGAAGCC	2520
50	CAGACATAGG ATGGGTCAGT GGGCATGGCT CTATTCCTAT TCTCAAACCA TGCCAGTGGC	2580
	AACCTGTGCT CAGTCTGAAG ACAATGGACC CACGTTAGGT GTGACACGTT CACATAACTG	2640
	TGCAGCACAT GCCGGGAGTG ATCAGTCRGA CATTTTAATT TGAACCACGT ATCTCTGGGT	2700
55	AGCTACAAAA TTCTCAGGG ATTTTCATTTT GCAGGCATGT CTCTGAGCTT CTATACCTGC	2760
	TCAAGGTCAK TGTCATCTTT GTGTTAGCT CATCGAAAGG TGTACCCTG GTTTCAATGA	2820
60	ACCTAACCTC ATTCTTTGTG TCTTCAGTGT TGGCTTGTTT TAGCTGATCC ATCTGTAACA	2880

CAGGAGGGAT CTTGGCTGA GGATTGTATT TCAGAACCAC CAACTGCTCT TGACAATTGT 2940  
 TAACCCGCTA GRCTCCTTTG GTTAGAGAAG CCACAGTCCT TCAGCCTCCA ATTGGTGTCA 3000  
 5 GTACTTAGGA AGACCACAGC TAGATGGACA AACAGCATTG GGAGGCCCTTA GCCCTGCTCC 3060  
 TCTCTATTCC ATCCTGTAGA GAACAGGAGT CAGGAGCCGC TGGCAGGAGA CAGCATGTCA 3120  
 10 CCCAGGACTC TGCGGGTGCA GAATATGAAC AAYGCCATGT TCTTGACAGAA AACGCTTAGC 3180  
 CTGAGTTTCA TAGGAGGTAA TCACCAGACA ACTGCAGAAAT GTRGARGACT GAGCAGGACA 3240  
 GCTGACCTGT CTCCTTCACA TAGTCCATRT CACCACAAAT CACACACAA AAACGAGARG 3300  
 15 AGATATTTTTG GGTTCAAAAA AAGTAAAAAG ATAATGTAGC TGCATTTCTT TACTTATTTT 3360  
 GARCCCCAAA TATTTCTTCA TCTTTTTGTT GTTGTCTATG ATGGTGGTGA CATGGACTTG 3420  
 TTTATAGAGG ACAGSTCAGC TGTCTGGCTC AGTGATCTAC ATTCTGAAGT TGTCTGAAAA 3480  
 20 TGTCTTCATG ATTAAATTCA GCTTAAACGT TTTGCCGGGA ACACTGCAGA GACAATGCTG 3540  
 TGAGTTTCCA ACCTYAGCCC ATCTGCGGGC AGAGAAGGTC TAGTTTGTCC ATCASCATTA 3600  
 25 TCATGATATC AGGACTGGTT ACTTGGTTAA GGAGGGGTCT AGGAGATCTG TCCCTTTTAG 3660  
 AGACACCTTA CTTATAATGA AGTATTTGGG AGGGTGGTTT TCAAAATTAG AAATGTCCTG 3720  
 TATTCCTATG ATCATCCTGT AAACATTTTA TCATTATTA ATCATCCCTG CCTGTGTCTA 3780  
 30 TTATTATATT CATATCTCTA CGCTGGAAAC TTTCTGCCTC AATGTTTACT GTGCCPTTGT 3840  
 TTTTGCTAGT GTGTGTGTTT GAAAAAATA ACATTCTCTG CCTGAGTTT AATTTTTGTC 3900  
 35 CAAAGTTATT TTAATCTATA CAATTAAAAG CTTTTCCTTA TCAAAAAA AAAAAAATA 3960  
 AAAAAAATA AAAAAGCGGA CCGTGGGC 3989

## (2) INFORMATION FOR SEQ ID NO: 29:

## (i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 3735 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

CTGCTGTTCG CTGGCTGGGC TCCGCAGCAG GCTTGGCCAG CSGCTGACGG GTCCGCGGGC 60  
 GGGTTTGTGT GAACAGGCAC GCAGCTGCAG ATTTTATTCT GGTAGTGCAN CCTCTCAA 120  
 55 GGTGAAGGA ACTGATGTAA CAGGGATTGA AGAAGTAGTA ATTCCAAAA AGAAAACTTG 180  
 GGATAAAGTA GCGGTTCTTC AGGCACTTGC ATCCACAGTA AACAGGGATA CCACAGCTGT 240  
 60 GCCTTATGTG TTCAAGATG ATCCTTACCT TATGCCAGCA TCATCTTTGG AATCTCGTTC 300

	ATTTTACTG GCAAAGAAAT CCGGGGAGAA TGTGGCCAAG TTTATTATTA ATTGATACCC	360
	CAAATATTTT CAGAAGGACA TAGCTGAACC TCATATACCG TGTTTAATGC CTGAGTACTT	420
5	TGAACCTCAG ATCAAAGACA TAAGTGAAGC CGCCTGAAG GAACGAATTG AGCTCAGAAA	480
	AGTCAAAGCC TCTGTGGACA TGTGTGATCA GCTTTTGCAA GCAGGAACCA CTGTGTCTCT	540
10	TGAAACAACA AATAGTCTCT TGGATTTWTT GTGTTACTAT GGTGACCAGG AGCCCTCAAC	600
	TGATTACCAT TTTCAACAAA CTGGACAGTC AGAAGCATTG GAAGAGGAAA ATGATGAGAC	660
	ATCTAGGAGG AAAGCTGGTC ATCAGTTTGG AGTTACATGG CGAGCAAAAA ACAACGCTGA	720
15	GAGAATCTTT TCTCTAATGC CAGAGAAAAA TGAACATTCC TATTGCACAA TGATCCGAGG	780
	AATGGTGAAG CACCGAGCTT ATGAGCAGGC ATTAACTTG TACACTGAGT TACTAAACAA	840
20	CAGACTCCAT GCTCATGTAT ACACATTTAA TGCATTGATT GAAGCAACAG TATGTGGCAT	900
	AAATGAGAAA TTTGAGGAAA AATGCAGTAA AATACTGGAG CTGCTAAGAC ACATGGTTGC	960
	ACAGAAGGTG AAACCAATC TTCAGACTTT TAATACCATT CTGAAATGTC TCCGAAGATT	1020
25	TCATGTGTTT GCAAGATCGC CAGCCTTACA GGTTTTACGT GAAATGAAAG CCATTGGAAT	1080
	AGAACCCTCG CTTGCAACAT ATCACCATAT TATTGCGCTG TTTGATCAAC CTGGAGACCC	1140
30	TTTAAAGAGA TCATCCTTCA TCATTTATGA TATAATGAAT GAATTAATGG GAAAGAGATT	1200
	TTCTCCAAAG GACCCGGATG ATGATAAGTT TTTTCAGTCA GCCATGAGCA TATGCTCATC	1260
	TCTCAGAGAT CTAGAACTTG CCTACCAAGT ACATGGCCTT TTA AAAACCG GAGACAACTG	1320
35	GAAATTCATT GGACCTGATC AACATCGTAA TTTCTATTAT TCCAAGTTCT TCGATTTGAT	1380
	TTGTCTAATG GAACAAATTG ATGTTACCTT GAAGTGGTAT GAGGACCTGA TACCTTCAGC	1440
40	CTACTTTCCC CACTCCCAA CAATGATACA TCTTCTCCAA GCATTGGATG TGGCCAAATCG	1500
	GCTAGAAGTG ATTCTTAAAA TTTGGAAAGA TAGTAAAGAA TATGGTCATA CTTTCCGCAG	1560
	TGACCTGAGA GAAGAGATCC TGATGCTCAT GGCAAGGGAC AAGCACCCAC CAGAGCTTCA	1620
45	GGTGGCATT TCTGACTGTG CTGCTGATAT CAAATCTCGG TATGAAAGCC AACCCTCAG	1680
	ACAGACTGCT CAGGATTGGC CAGCCAGCTC TCTCAACTGT ATAGCTATCC TCTTTTAAAG	1740
50	GGCTGGGAGA ACTCAGGAAG CCTGGAAAAT GTTGGGGCTT TTCAGGAAGC ATAATAAGAT	1800
	TCCTAGAAGT GAGTTGCTGA ATGAGCTTAT GGACAGTGCA AAAGTGTCTA ACAGCCCTTC	1860
	CCAGGCCATT GAAGTAGTAG AGCTGGCAAG TGCCCTCAGC TTACCTATTT GTGAGGGCCT	1920
55	CACCCAGAGA GTAATGAGTG ATTTTGCAAT CAACCAGGAA CAAAAGGAAG CCCTAAGTAA	1980
	TCTAACTCCA TTGACCAAGT ACAGTGATAC TGACAGCAGC AGTGACAGCG ACAGTGACAC	2040
60	CAGTGAAGGC AAATGAAAGT GGAGATTGAG GAGCAGCAAT GGTCTCACCA TAGCTGCTGG	2100



	AATCACACCT GAGAACTGAG ATATACCAAT ATTAAACATT GTTACAAAGA AGAAAAGATA	2160
	CAGATTTGGT GAATTTGTTA CTGTGAGGTA CAGTCAGTAC ACAGCTGACT TATGTAGATT	2220
5	TAAGCTGCTA ATATGCTACT TAACCATCTA TTAATGCACC ATTAAAGGCT TAGCATTTAA	2280
	GTAGCAACAT TGGCGTTTTC AGACACATGG TGAGGTCCAT GGCTCTTGTC ATCAGGATAA	2340
10	GCCTGCACAC CTAGAGTCTC GGTGAGCTGA CCTCAGCATG CTGTCTCTCT GCGATTGCCC	2400
	TCTCTGCTG CTGGACTTCT GCTTTGTGTG GCCTGATGTG CTGCTGTGAT GCTGGTCTTT	2460
	CATCTTAGGT GTTCATGCAG TTCTAACACA GTTGGGGTTG GGTCAATAGT TTCCCAATTT	2520
15	CAGGATATTT CGATGTCAGA AATAACGCAT CTTAGGAATG ACTAAACAAG ATAATGGCAG	2580
	TTTACGCTGC ACAACTGGTA AAATGACTGT AGATAAATGT TGTAAATAGT GTACACGTTT	2640
20	GTATTTTTGT TAATATAGCC GCTGCCATAG TTTTCTAACT TGAACAGCCA TGAATGTTTC	2700
	ATGTCTCCCT TTTTTTTTTG TCTATAGCTG TTACCTATTT TAGTGGTTGA AATGAGAGCT	2760
	AGTGATGACA GAAGGATGTG GAATGTCTTC TTGACATCAT TGTGTATTGC TGGTAATCAA	2820
25	GTTGGTAACG ACTACTTCTA GCAGCTCTTA CCACTATGAC TTAAGTGGTC CTGGAAGGCA	2880
	GTAAGTGGAG GTTTGCAGCA TTCTTGCCTT CATGAGGGCT TCTACCACTG ACCACTTTGC	2940
30	ACGTACCTGG CTCCGAGATT TACTTAGGTA CCCCACGAGT CGTCCACATA AGCAGCTTCA	3000
	TCTTTACCTT GCCAGAGTTG ACAATTATGG GATACTCTAG TCTACTTATA CTTGTGTTC	3060
	CATCTGTCTG CCATCCTCTG AAGGCCAGGA CCCAGTCATA CATCCTTAGA AACCAAAGTA	3120
35	TGGTTTTTGT TTTCTCTTGG AATGTCAGGT CTTAAGGCAT TTAATTGAGG GACAAAAAAA	3180
	AAAAAAGCC GATATAGTAG CTAGCTACTT AAGCATCCAT GGGTATTGCT CCATATCAAA	3240
40	GCAGATTGCG AGGACAGAAA GAGTAAATTA GCCTTCAGTC TTGGTTTACA GCTTCCAAGG	3300
	AGAGCCTTGG CCACCTGAAA TGTTAACTCG GTCCCTTCCT GTCTCTAGTT CATCAGCACC	3360
	TGCAGATGCC TGACTCTTGT TAGCCTTACT ATTCAATACA GTCCCTTAGAT TCACCGTATG	3420
45	CCTCTTCCTA TCCAGGCACC TATCTGAAT CACCATGTTG CTCTGCAGCT AGAGTTGATA	3480
	GGAGAAAATC CATTTGGGTA GATGCCCTAT GAATTTGTAG TAGACTTTCA AAATGAGTGA	3540
50	TTTGTTAGCT TGGTACTTTT AAGTTTGTGG TACAGATCCT CCAAACCCAT ACTCTGAGCA	3600
	ATTAAGTCCC TTGAACATAG AGAAAATTAA GGCTCACAG GATGAGTCTC CATCTCTGT	3660
	AAATGCTTAT TTTATCATAG TCTTTAGCCN CTACTATGAG TAAATGTTT TCTTCNGCCG	3720
55	GGTGTGGTGA CTCAC	3735
60		

## (2) INFORMATION FOR SEQ ID NO: 30:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1667 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

10 TAGTAATTCA TTAACTCCT CTTACATGAG TAGCCACAAT GAGTCAGATA TCGAAGATGA 60  
AGACTTAAAG TTAGAGCTGC GAGGACTACG AGATTAACAT CTCAAAGAGA TTCAGGACCT 120  
15 GCAGAGTCGC CAGAAGCATG AAATTGAATC TTTGTATACC AACTGGGCA AGGTGCCCCC 180  
TGCTGTATT ATTCCCCAG CTGCTCCCT TTCAGGGAGA AGACGACGAC CCACTAAAAG 240  
CAAAGGCAGC AAATCTAGTC GAAGCAGTTC CTTGGGAAT AAAAGCCCC AGCTTTCAGG 300  
20 TAACCTGTCT GGTGAGAGTG CAGCTTCAGT CTTGCACCCC CAGCAGACCC TCCACCTCC 360  
TGGCAACATC CCAGAGTCCG GGCAGAATCA GCTGTTACAG CCCCTTAAGC CATCTCCCTC 420  
25 CAGTGACAAC CTCTATTGAG CCTTCACCAG TGATGGTGCC ATTTCAGTAC CAAGCCTTTC 480  
TGCTCCAGGT CAAGGAACCA GCAGCACAAA CACTGTTGGG GCAACAGTGA ACAGCCAAGC 540  
CGCCCAAGCT CAGCCTCCTG CCATGACGTC CAGCAGGAAG GGCACATTCA CAGATGACTT 600  
30 GCACAAGTTG GTAGACAATT GGGCCCGAGA TGCCATGAAT CTCTCAGGCA GGAGAGGAAG 660  
CAAAGGCAC ATGAATTATG AGGGCCCTGG AATGGCAAGG AAGTTCTCTG CACCTGGGCA 720  
35 ACTGTGCATC TCCATGACCT CGAACCTGGG TGGCTCTGCC CCCATCTCTG CAGCATCAGC 780  
TACCTCTCTA GGTCACTTCA CCAAGTCTAT GTGCCCCCA CAGCAGTATG GCTTTCAGC 840  
TACCCGATTT GGCCTCAAT GGAGTGGGAC GGGTGGCCA GCACCACAGC CACTTGGCCA 900  
40 GTTCCAACCT GTGGGAAGTG CCTCCTTGA GAATTTCAC ATCAGCAATT TGCAGAAATC 960  
CATCAGCAAC CCCCAGGCT CCAACCTGCG GACCACTTAG ACCTAGAGAC ATTAAGTGA 1020  
45 TAGATCTGGG GGCAGGAGAT GGAATGCTGA GGGGTGGGT GGGGTGGGA AGTAGCCTAT 1080  
ATACTAACTA CTAGTGCTGC ATTTAACTGG TTATTCTTG CCAGAGGGA ATGTTTTTAA 1140  
TACTGCATTG AGCCCTCAGA ATGGAGAGTC TCCCCGCTC CAGTTATTGG AATGGGAGAG 1200  
50 GAAGGAAGA ACAGCTTTTT TGTCAAGGGG CAGCTTCAGA CCATGCTTTC CTGTTTATCT 1260  
ATACTCAGTA ATGAGGATGA GGGCTAGGAA AGTCTTGTTC ATAAGGAAGC TGGAGAACTC 1320  
55 AATGTAAAT CAAACCCATC TGTAATTTG AGTGGGTGGA GCTCTTGCTT TTGGTACATG 1380  
CCCTGAATCC CTCACTCCTT CAGAATCCG AACCACAGGA CAAAACCAC CTACTGGGCT 1440  
CTCTCCTACC CTGCCCTCCT CCTTTTTTT TACCCCTCTC TTTTTATTT TTTCTTGCT 1500  
60

CTTTAGAACT CAGTGAAAAA TACCAGGGTA CTGGGGTGCA ACTCTTCTT ATGATAGGTC 1560  
 ATTAGTGCTT TAAGCAAAAG ATATTAGCAG CTTTGACTGC AGCATTAGCA ATTAGGFAAA 1620  
 5 AAAAAAANWA AAAACTCGAG GGGGGGCCCC GTTACCCAT TCGCCCT 1667

10 (2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1408 base pairs

(B) TYPE: nucleic acid

15 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

20 ATTACACACC TGAGCACTGT GCCTGGCAAG ACCTGTCTTA ATAGATTAGA GAACCACTGA 60  
 TAGATGGTCA GCTTTCTGTA GCAGTGAGAA CCTACATTT CAAATGTGGA TAGCACCTTT 120  
 GCGGGGAAAC ATCACTTGGC ACATCTGCAT TCTTTTTTGA CACAGGTCT CACTCTGTTG 180  
 25 CCCAGGCTAG AGTGCATGGC ACGATCTTAG CTCACTGCAA CCTCCACCTC CCAAGTTCAA 240  
 GCGATTCTTC TGCTCAGCC TCCTGAGCAG CTGGGATCAC AGACATGGC TACCATGCCC 300  
 30 AGCTAATTTT TTGTATTTT TGTATGTTG TTTTGTITK TAAGTAGAGA CGGGCTTTCA 360  
 CCACGTTGGS CAGGCAGGTC TCGAACTCCT GAMCTCAGG GATCCACCCA CATCTCGGTT 420  
 CCAATATCTT TCTCAACATA ATGATAGCCG TAATTAATAT TTTCCAGTAC ATTTTATGC 480  
 35 CTTTACACAC GAGAGTGGTA GACAGACACA AACCCAGATC TGTCTGACTC CAAAGCCCGT 540  
 TTGTCATCAT TCCTTTTACG GTATCCTATA GTGGTATCCT TTACAGAAAG ACAGCTTTTA 600  
 40 CCCAACAAAG ACTTAACTTC CCAGGATGCC AGAAGGACAA AGCGGGATTG CTTTAAAGRA 660  
 GRAAGTATC AAGAMCTTAT TTTATAAATG AGATTAGATA GGGAAAGGCA ATTTATCTTT 720  
 ATTAAAAACT GAAAGGCCA GCATAGGGAA GGAGTCCTT CGGTGGTCTT TTTACGGGAA 780  
 45 ATACTTCAGT TGCTTTTATT AGAAACAGAT AGTACCTAAG GTTTGAGGT AGGWACAGCT 840  
 TAAGGCATGC TAATGKTCAT GGGTCCTTCC ATAGTCATTT TGTATTTTG GTTWACATTT 900  
 50 GAGCAATAGG CAGCCCTTCA CTGCTGCTGG ATCATTCTCT GCCATATTA CAGGTGACAG 960  
 AGGAGACAGG AGGTATGTCT TTTCTATTTT TAWACATGCT TTATATTTAA CACAAGCTCT 1020  
 TGGGTATCTT AGATAAACAG AAGTTGCCTA GCACTCCTTT TAGTGCAATT AACCTTTAA 1080  
 55 CATTTAAGCA AAATAATAAA CAGTCTTTTG AGGTTCCCTA ACAATGAAAC GTGTTCCAGT 1140  
 GGCAGCAGCG GAATCCATGC YTCCTCTCCT GGAGTGTGCA AKAGTCCGTG GTCCTGAGTA 1200  
 60 TCTCACACAG ATGTGGCATT TTTGTGTGTA TGCTCTAATT AAGGCCATG GTACAGAACC 1260

AGATTGAGAC GTCTCTCAG AATTAATGCA TTCTTTTGCA AAGGTGAATA TTTTCTCTT 1320

AAAAAATATG TATAGGTGG TATGTCATT TATTAGTCTT GCTAAAAAA AAAAATAAA 1380

ACTTNGAGGG GGGGACGGT ACCGATT 1408

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2031 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

AGGATATGCA TGATCTTAA CCGGGTATA TGTTAAAAA AAATTGGAAA ATGCAATACA 60

TTTTTACTA TACAACTAC AGATGAGTA TGCAAGTTT ATTTATCAA ATGTAATGGA 120

TTTTTAAGG CTGAGAAAT TTCTTATAC CTACCTTTC AGTTATTTA ATTATACCA 180

ATTATCACT AGATAGCTT CATCCATAG AATATAAAA TGAAGAGACA CCTACGCTT 240

ATCAGGCTTA GGAATCTTG AACTTATTC CACTTTAAT TCTCAGTGA AGTTAAGAG 300

GGTGAGAAA CAAGAGGGG GAATACTGA CACTAACAA AACCGGCACC ACATCGCTAG 360

GTGGTGCTA CTAATTAAT TCTCAGGAT TTCTCAGAT TGAAAAGCTT ATGAGGATTT 420

CTTGGGATC TTAATAAAT GCTGTAGT ACAGAGCTT CCGATGATA TTTACTCTG 480

AGCATGCTG GTTGTAATC CTAACTTTC TTTCTCCAG AGGGTGGTGA TAGAAACAGA 540

TGGTAGTAT TAGGAATGA TGTCTCGTG AAATGTTGAG GGTGGGGAGA AAGACTTTA 600

AGGGAGGAGA GCCATCTAT TTGTCCTAA AGCCACCTT CAGCAGAATC GTCATGTTTT 660

TCTGATGCAC CGCTCTGCTT CATGCCAAG ATGACTTGG AGGCAATCTC AGGAGCTGTG 720

CACTTAACCT TTGCAAGCA CACTGTCTT CTCAGCGTC TCTGCAAGTC AGTAGGTGT 780

AGTATGGTTG CAAAGTTCAC TGTCTCAGCA AAGTTGAAT GGGCTACCTC TCTACAGCTG 840

TTTCTCTGA GGGAAAAATC TTGAGACCAG ATGGTGGAGC TCTGGAGTCA CAGGAAATGG 900

GTGCTCTCAG CACAAGCTG CTGCTTTTAC TTCAGCCACT TCTGACATTT TTACATACCG 960

AGCCTGAGAT TGTGTGATTA TCTCAAATCA AATCACTTTG ATGGAGATAA ATATCAAAA 1020

CTGTTTTATA GTCATTGATT TGGTGAGAAC AGTAATGGAA AATGGTGTG AAGGACTTCT 1080

CATTTTTGGA GCTTTCCTTC CAGAGTCTG GCTGATGGT GTTCGCTGTT CATCTGAGCC 1140

CCCAAAAGCA TTATTACTGA TACTTGACA CAGTCAAAAG CGCAGACTGG ATGGATGGTC 1200

TTTTATAAGG CATTTAAGGG TACACTACTG TGTTPCACTG ACCATACATT TTTCTTAGCC 1260  
 CCTCAAGTAA TATAGCACAG AGTTATGAAT GACAATTCCC CTAACCATTC CTCTTCATAT 1320  
 5 CTGCCCTCTC CCCTTACCAT CGTAATTCTC CAACTGGTC ATAAAGGCAC TCTGTGAAGA 1380  
 TATTGGGGAC TGACATCTTA AGCTCTCACC TGGCTGCAGT AGGAAAGGCC AACTGACGA 1440  
 CAAAAAATA ATTCTTTATA AAGATGATAT GGTAAATGT ATCTTTGCCC TGGGTCTGGG 1500  
 10 TGGGTCCAGT CAGTCTCAGA TTTACAAGCA TTTAGGAGCC TAGGTAAAAG CTGCTAGTAT 1560  
 TCTTTTAAAA GTTACATTTA TGACTTGCAA TCATAGAAA CTCCTTCCAA TTAAATGGCA 1620  
 15 TTTTATAATA TTATGTGTGT ACTTCACAGT GTTAAAAATA CCCTCATACG TTATTGCATT 1680  
 TGATCTTCAC AGAAAGTGCA TTTTAACCAG TACTCTGGGT GCAATAAATA ATATGTAGAA 1740  
 ATTTAAGTCC TCCAATTCCA GCATATCCAG TGAGTTTGA CAGTGTGTTT ATGTGGAATG 1800  
 20 TTTAAGGATA TACAATTGTA CTTTATATAA ATTGGTCTT GTTCTTCTTA AATGTGACAT 1860  
 GAAATAATTG TCCTGCTACA TTATACTGGA AATTAAACAG CGAAAAGGGA AGAGCTCTTG 1920  
 25 GCTCCCTTGA GGTCTGCTA GTGGTGTTAG GAGTGGTTAC AACTGAGCTT TTAGTAACCA 1980  
 TTTAACCCTA TGTAACCTG GTTCTAATT AAAAAAAT TTCTTTTCC A 2031

30

(2) INFORMATION FOR SEQ ID NO: 33:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 971 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CGCGTCGGAA CTCGGCCGCG GCACATCCAC GGGGCGGAG TGACACCGCG GAGGGAGAGC 60  
 AGTGTCTGCG TGGAGCCGAT GCCAAAACC ATGCATTCT TATTGAGATT CATTGTTTTT 120  
 45 TTTTATCTGT GGGCCCTTTT TACTGCTCAG AGACAAAAGA AAGAGGAGAG CACCGAAGAA 180  
 GTGAAAATAG AAGTTTGGCA TCGTCCAGAA AACTGCTCTA AGACAAGCAA GAAGGGAGAC 240  
 50 CTACTAAATG CCCATTATGA CGGTACCTG GCTAAAGACG GCTCGAATTT CTACTGCAGC 300  
 CGGACACAAA ATGAAGGCCA CCCCAATGG TTTGTTCTTG GTGTTGGGCA AGTCATAAAA 360  
 GGCTAGACA TIGCTATGAC AGATATGTGC CCTGGAGAAA AGCGAAAAGT AGTTATACCC 420  
 55 CTTTCATTTG CATACGGAAA GGAAGGCTAT GCAGAAGCCA AGATTCCACC CGATGCTACA 480  
 TTGATTTTTG AGATTGAACT TTATGCTGTG ACCAAAGGAC CACGGAGCAT TGAGACATTT 540  
 60 AAACAAATAG ACATGGACAA TGACAGGCAG CTCTCTAAG CCGAGATRAA CCTCTACTTG 600

CAAAGGGAAT TTGAAAAAGA TGAGAAGCCA CGTGACAAGT CATATCAGGA TGCAGTTTTA 660  
 GAAGATATTT TTAAGAAGAA TGACCATGAT GGTGATGGCT TCATTTCTCC CAAGGAATAC 720  
 5 AATGTATACC AACACGATGA ACTATAGCAT ATTTGTATTT CTACTTTTTT TTTTATAGCTA 780  
 TTTACTGTAC TTTATGTATA AAACAAAGTC ACTTTTCTCC AAGTTGTATT TGCTATTTTT 840  
 10 CCCCTATGAG AAGATATTTT GATCTCCCCA ATACATTGAT TTTGGTATAA TAAATGTGAG 900  
 GCTGTTTTGC ARACTTAAA AAAAWWAAA AAAACTSGAG GGGGGCCCGT ACCCAANTCG 960  
 CCGNATATGA T 971

20 (2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1792 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GAACCCCTT TCTCCTGGTA AAGGGTAAGG GGGGGGATAA TGTTTACCAC AGGTACGAAA 60  
 30 TAGTCACTTT AACATTGAGA CCTCTGCCTC ATTGAATTCA GGTTTTTTAA GTACTTGAAA 120  
 CTCTTCAGAT TCTCCTTATT TTAGTTTCTT TTTACATTTA TGAAGTAGAA AGCATTGTTT 180  
 35 TGTAAGTGT TTTGAAAATA AATAGCCTAG TCTCTTATCC TCTTTAGCGT GGATTAAAGG 240  
 TGAAGTTCTG CAAATGGGAG AGTGTTTACA GTAGATAGCT CAGATTGATT GAACACATTT 300  
 GAGGAAGAGA CTCCTGCATG AGATACCAGC ATTTTACAA ATACTTTTTA TGTACATTCT 360  
 40 TTATTTTGTG ATTTTGTCAA CCTCTCCCC AAGCACATCT TCTTTCCTTT TACTATGTCT 420  
 ATGTAGGGAA AAACAAAACA AAAAATTGCA CTTACGTTAC ACTCCCAAAA TGTGGGTAAT 480  
 45 CCGTGTCTTT CAAAAACAT TTCTGTTTTT TGTMTTGMTT TGGTCAGTCC ATTGCATAAG 540  
 TGACAAGTTT GGGTGCTTGT GGCACGTATG TATGAAGCGG GAGGGGGATG ASAATTGCCT 600  
 GTCCTTCAGT ARGCTGTAAA AGTAATTTAC ATGTAAGTAA AAAGGGAAAA TAGAATAGAT 660  
 50 GCCAAAGTCA TTTATTCAGT CCTTAGTTTT CTTATGTGGC ATTACTGCAT CTGCTAGTTA 720  
 GTGAGAAAGC ACCCTCAGCT TTTACTGCTC CCCTGCCTGC CTGCCAACAC ACTTGATGTG 780  
 55 TGCAACAGC CCTCAAGTAT CTGTCTGATG ACCTATATAA GGTATTGAAT AAGGTATTCT 840  
 TGTCAGTTTA GAAATGGACT GGATAAACT TACTTGGTTG TCATTATTTT ATCTCATTG 900  
 TCCTGTTACA TGCCCTATGT TAAGATAATT ATATTGCCAC TAATAATCAA GATCCTAAAT 960  
 60

GAGTATTACA ACTGGCTAAT ATCATTTTTT ATATACAAGG GTATGTGTAT ATTTGGAATT 1020  
 GGTATGAGAA ACTCATTTGT ACCCATTTGA GTGATATTGC ACAACAAACA CAGATAYCTA 1080  
 5 CAGACTCCGT TTTCATTTTC TCGTGTCTT TATGATAATG ATCTTTGTAG ATTGCTTATT 1140  
 TCTGTACTTT ATCTGTAATA AACTTTGTAG ATCCTGTGAA CCATTACTTT GCCTAAATCA 1200  
 10 CTTGAGACTT GAGTCTTTAA TAACAAAGCA TCAATATTCA CTAAAGTCAA TCTCTTTTGA 1260  
 GTTCTGTGA CTTGGCTAGA AGCTCTTGAC ACTAAGGGAT TAGTGTTAAT TTCCCTGGG 1320  
 GGTGTCCAC TAGGGCATT CTGTATAATG ACTTGATGTT GCCACATAGA CTTCAAGATA 1380  
 15 TATAATATTT TGAGGATTTT GTTGATGGC CTATGTTTTA TGCATAGTG TGAAACGTGT 1440  
 AAAGCTTGGT TAACCTGTAT ATAGATAGCT TATTGTTGAC TAGTTATAGT GTATTTAGGG 1500  
 20 TTGCTGTAA TATTTAAGCT TCTTTACTGA TGTGTGTGCT GGTAGGAACA TATAATTTTT 1560  
 GTACATTATA TTTACTGAGA TGTGCTTTT TTTATTTTAC AAATACTTTG GAATCCCAAT 1620  
 GTGTTTTTTG CTTCCGTGAG GATTAATTTG GAAAGGTTTT TAATGACATT CCACTGATTT 1680  
 25 CAGATTTTGC TTGAGATTGA CTTCAATAAA TTGTCCTGTA TGTTCACAAA AAAAATTAAA 1740  
 AAACTCGAGG GGGGCCCCGT ACCCAANNCG CCGGATATGA TCGTAAACAA TC 1792

30

(2) INFORMATION FOR SEQ ID NO: 35:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 896 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

AGTTGNANAC AACAGGACCT GAGTCCTTGG GCAGCACCAG TAGGTTGCCC CYTGCTCTT 60  
 GCCAGCTCA CYTGCCACTT TYTGCCCTTY TCGGGATGCC TTCCGAGACA GAGTYTTTCG 120  
 45 CTGCTGTGG TGGCCAYTCT TTGCTTTTGG TTYTCTTGCC CCTTGGCTC CCTTTTGTG 180  
 CCGGGGAGC CTTGTGTGAC CTGCCCTTTT CCCTCCCTTC CTTTCAGGA CAAGCAGGCC 240  
 50 GAGGAGGTGC CGAAAAACAA GGAGCTGAAG GAAGAGGCCT CCAGGTAAG CCTAGAGGCC 300  
 AAAGAACCTT CCAGGTGAGC CGGACAGCTC CAGCAGCTCC ACGTTCAGG CAGCCTCGMC 360  
 CGCCGGCTGC GCTCCACGA CTGGGTTTGG GGGGAGGGG GGTGGCCAAG GGGCGTTTCC 420  
 55 TCTGCTTTTG GTGTTGTAC ATGTTAAGAA TTGACCAGTG AAGCCATCCT ATTTGTTTCC 480  
 GGGGAACAAT GACCGGCTCG CAGAGGGAG AGGAGAGAGT TTGGGAAGG GAGATGGAGA 540  
 60 AGAACTCAAG GACATTCCAA CCTGCCCCG CCGAGATCTG ATTTTCACAT CTCTACCTGG 600

ACATTGAGCC TCCCAGGCAC CATGTTGAGG AGAGATGAAA ACCAGGGCGG TAGAACTTCA 660  
 5 GGGTGAAGCA CAGAGTCCTG GGTGGGGCAG CGGTGCGAGG GCGCACCAGA GAACCCAGCC 720  
 AGAGGGGGTG TGAGTACCAG TGGTGTGCT TCCACCCTGC AGCAGGTGGG ATGAGGTCTG 780  
 TGTGTGTGTG TGAACCATCA TTTTGTGATC ATCATGACCA ATGAAACATT GAAAAAATAA 840  
 10 AAAAAAATG GAGGGGGGCC CGTACCCAAN TCGCCGNATA GTGATCGTAA ACAATC 896

15 (2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 912 base pairs  
 (B) TYPE: nucleic acid  
 20 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

25 TCGACCCACG CGTCCGGTCA GCCAGTCGCA TCCAGCCATG ACAGCCTTCT GCTCCCTGCT 60  
 CCTGCAAGCG CAGAGCCTCC TACCCAGGAC CATGGCAGCC CCCCAGGACA GCCTCAGACC 120  
 AGGGGAGGAA GACGAAGGGA TGCAGCTGCT ACAGACAAAG GACTCCATGG CCAAGGGAGC 180  
 30 TAGGCCCCGG GCCACCGCGG GCAGGGCTCG CTGGGGTCTG GCCTACACGC TGCTGCACAA 240  
 CCCAACCCTG CAGGTCTTCC GCAAGACGGC CCTGTTGGGT GCCAATGGTG CCCAGCCCTG 300  
 35 ARGGCAGGGA AKGTCAACCC ACCTGCCCAT CTGTGCTGAG GCATGTTCTT GCCTACCATC 360  
 CTCCTCCCTC CCGGGCTCTC CTCCCAGCAT CACACCAGCC ATGCAGCCAG CAGGTCTCTC 420  
 GGATCACYGT GCTTKGGTGG AGGTCTGTCT GCACTGGGAG CCTCARGAG GCTCTGCTCC 480  
 40 ACCCACTTGG CTATGGGAGA GCCAGCAGGG GTTCTGGAGA AAAAAACTGG TGGGTTAGGG 540  
 CCTTGGTCCA GGAGCCAGTT GAGCCAGGGC AGCCACATCC AGGCGTCTCC CTACCCCTGGC 600  
 45 TCTGCCATCA GCCTTGAAGG GCCTCGATGA AGCCTTCTCT GGAACCACTC CAGCCCAGCT 660  
 CCACCTCAGC CTGGCCTTC ACGCTGTGGA AGCAGCCAAG GCACTTCCTC ACCCCCTCAG 720  
 CGCCACCGAC CT/TTTGGGG AGTGGCCGGA AAGCTCCCSG GCCTYTGGCC TGCAGGGCAG 780  
 50 CCCAAGTCAT GACTCAGACC AGTCCCAGA CTGAGCTGCC CACACTCGAG AGCCAGATAT 840  
 TTTTGTAGTT TTTATKCTT TGGCTATTAT GAAAGAGGTT AGTGTGTTCC CTGCAATAAA 900  
 55 CTGTTCCTG AG 912

60 (2) INFORMATION FOR SEQ ID NO: 37:



## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1382 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

10	AATTCGGCAC GAGCGGAGGC GAGGCAAACT RAGGGCGAAA GTTGTGTGTC GTGTTGGCAG	60
	GAGGGCCTAG AAGCGAAAGA CTGTCTAGTG GGACAATGTC ATATTATAAA TTTGGAATGC	120
	TGAATAGAAA ATTATAGATT TTGATATTGA AGGAAATGAA GCGAAGCTTA AATGAAAATT	180
15	CAGCTCGAAG TACAGCAGGC TGTTCGCTG TTCGTTGTT CAATCAGAAA AAGAGGAACA	240
	GACAGCCATT AACTTCTTAT CCACTTAAAG ATGATTCCAG TATCAGTACC CCTTCTGACA	300
20	ATTATGATTT TCCTCCTCTA CCTACAGATT GGGCCTGGGA AGCTGTGAAT CCAGAGTTKG	360
	CTCCTGTAAT GAAAACAGTG GACACCGGGC AAATACCACA TTCAGTTTCT CGTCCTCTGA	420
	GAAGTCAAGA TTCTGTCTTT AACTCTATTC AATCAATAC TGGAGAAGC CAGGGTGGTT	480
25	GGAGCTACAG AGATGGTAAC AAAAATACCA GCTTGAAAAC TTGGRATAAA AATGATTTTA	540
	AGCCTCAATG TAAACGAACA AACTTAGTGG CAAATGATGG AAAAAATTCT TGTCCAATGA	600
30	GTCGGGGAGC TCAACAACAA AAACAATTAA GAACACCTGA ACCTCCTAAC TTATCTCGCA	660
	ACAAAGAAAC CGAGCTACTC AGACAAACAC ATTCATCAAA AATATCTGGC TGCACAATGA	720
	GAGGGCTAGA CAAAAACAGT GCACTACAGA CACTTAAGCC CAATTTTCAA CAAAATCAAT	780
35	ATAGANACA AATGTTGGAT GATATTCCAG AAGACAACAC CCTGAAGGAA ACCTCAATTGT	840
	ATCAGTTACA GTTTAAGGAA AAAGCTAGTT CTTTAAGAAT TATTTCTGCA GTTATTGAAA	900
40	GCATGAAGTA TTGGCGTGAA CATGCACAGA AACTGTACT TCTTTTGA GTATTAGCTG	960
	TTCTTGATTC AGCTGTTACA CTTGGCCCAT ATTATTCGAA GACTTTTCTT ATGAGGGATG	1020
	GGAAAAATAC TCTGCCTTGT GTCTTTTATG AAATCGATCG TGAACCTCCG AGACTGATTA	1080
45	GAGGCCGAGT TCATAGATGT GTTGCAACT ATGACCAGAA AAAGAACATT TTCCAATGTG	1140
	TTTCTGTCAG ACCGGCTCT GTTTCTGAGC AAAAACTTT CCAGGCATTT GTCAAAATTG	1200
50	CAGATGTTGA GATGCAGTAT TATATTAATG TGATGAATGA AACTTAAGTA GTGATAAAG	1260
	GAAGTTTAGC ATAAATTATA GCAGTTTCT GTTATTGCTT AATTACCAT CTCCATAGTT	1320
55	TTATAGCTAC TATGTATTT CACTTGTGA ATTAAGTAT TTGAATTCTT TAAAAA	1380
	AA	1382

## (2) INFORMATION FOR SEQ ID NO: 38:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 872 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GGGCTACTTC AAAGCCCTGG GCCTTATTTT TTCAGGTAAA AAAATATAAA GTCAGATCTC 60  
ATCCCGGGCTG GCCATGCTGT TAGACCCCTT CATCCTTCTC TTCTGCCTCT TCTCAACAGC 120  
TGCCCACTCC TGTTTGGAAT TCATATACAT ACAGTTCTAA TACTGATGTA TTTACCCTCA 180  
TAAGCCACTC AACCAGAAT CTTATTTGAA TTATAATCCA GAAACATCAG GTGACGTGTG 240  
AGACTACTGT ATGAGAAAGA GACAGTTTAA GGGTCAGTCC AATGGAAAAA AGAGTTCTCA 300  
GAGCTTTCTT TAGCTTATTC TCATCAAAGA GCTTTCTCTG CAGAAGGAAC CTACTGGTTC 360  
CTCCTTTCCA GTCCTAGAAA TCCTGACCTA GAGTGGCTTA ATCCTGCTAG CACCTCTCTC 420  
TCGCACTCTG GTGCCAAATG ACTCCAGGAA CTGGGCCATG ATGTGGTGGG AATGACCTTA 480  
CCCTGAGCAT GTCACTCATG CATTGAACAA CAGCTAAGAG CAGAGCTTAG AGCTTAGAGC 540  
TGGGCCCTGT AAGGTGAGAG GAATCACATC CTGCAGAAGT CTGCTCTGAG AAGCAGGTAC 600  
TCCTGTCACA GCAGAGACAC AGTGGATACC TGAGTAACAA TAATACAAGA CAGGACGTGG 660  
GMACAGCAAA AGATTGGGT GTCAGAAGAR GCCGAGAACA CTTYCAGGCA GGAACATTCA 720  
RARTGTCTT TGGAGGAART AGGCMCSAAG GCTGGGCAGG ATTCMCGGG GCAGAGATGG 780  
AGCAAGCAAT TGAATGAAA GCCATGGCAT GGGAAAAGGA GCACTGGCCA CAGGGAGTGC 840  
AACGTTGTGA TGCAAGGCCA CTGTGGAGCC AT 872

## (2) INFORMATION FOR SEQ ID NO: 39:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 812 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

GGCAGAGGCT CACCCAGCA GAGATTGAGG GGAACCGTG ATGAAATTTT TAAGTATTCT 60  
GCTTGATGAT AATAATTTTY CTCTTATGTT AATGTTGGCT CCGTTTGGGT GTTTAGCTTT 120  
TGAAAGGAGT ATGAAATGC GGAATGGGGC TTTGGGGCTT GAGGAGGTGT GATCTCTAGT 180  
GTTTAAAAAA TTTAATTGCA CAAATAGAAA TAATTCACCC ACATTATTGA ACCCCACTAA 240

AGCATATCCT TTTTGTCAT ATTCTTTTCC TGCTGCCCTC GTGTGTACCA TTATTACTCA 300  
 GTTGTGATTT GAGCTCGTTC CACTTAAAGT CATTATAGA TACTTTTGGG TCGTGTTKGA 360  
 5 ATATTTATTC AATTTCTATT CTGTGTTTTA CTTAATTACT TTATTATGGA ACCTTTACAC 420  
 AGGTCTGGTG TACTTGTICT TTGAAAAGTC TTATGTTGAC CACCATCACT GAGCATATAG 480  
 10 CTTTTCCTT ATTTCTTGG GATAATTACC CGAAGTGGA ATACCGAATC AAACCTCTGT 540  
 TTTCTTCTT TGGCACTATT ATATAAATTG TTTTCCAAAC AAGGCATGTT TACAATAGAC 600  
 15 ATTTTTCAAA ATCTGGGTAT TTGCTCTATT TTGCTCTCTG TATGCAGAAT TCAGCGGGGT 660  
 GCCAAGTCGT TTTCTGTGTG GGTTGAGAGA CAGGCTGTGC AGCCCACTGT TGCATAGGAC 720  
 TAACTACTAC AAATCATGCT GAGACCGAGC TATTTTGGCT GCTTAGARGC TTTGCAGCCT 780  
 20 TGAGTAAGTT TCGNCATCTG GAAACNITGN AA 812

25 (2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1515 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

35 AATTCGGCAC GAGGGAAATT CAAGCACTTT TCCTAAAAGA AGGGGGAATG GATGCTGAAA 60  
 CAACACGTNT CCCACAAAGG GAGCAGACAC TGGGCTTGTG AAGCTGCCCC ATACCTTCCC 120  
 CACAGAACTG GGGTCCGGCC TCCCTGACAT GCAGATTTC ACCCAGAAGA CAGACAAGGA 180  
 40 GCCAGTGGTC ATGGAATGGG CTGGGGTCAA AACTGCGGTG CCTGGGAGCT GAGGCAGCCA 240  
 CCGTTTCAGC CTGGCCAGCC CTCTGGACCC CGAGGTGGA CCCTACTGTG ACACACCTAC 300  
 45 CATCGGGACA CTCTCAACC TCCTCTGGCT TGCCTGGCC TGCAGCCCTG TTCACACTAC 360  
 CCTGTCAAAG TCAGATGCCA AAAAAGCCGC CTCAAAGACG CTGCTGGAGA AGAGTCAGTT 420  
 TTCAGATAAG CCGGTGCAAG ACCGGGGTTT GGTGGTGACG GACCTCAAAG CTGACAGTGT 480  
 50 GGTCTTGAG CATCGCAGCT ACTGCTCGGC AAAGGCCCGG GACAGACACT TTGCTGGGGA 540  
 TGTACTGGGC TATGTCACTC CATGGAACAG CCATGGCTAC GATGTCACCA AGGTCTTTGG 600  
 55 GAGCAAGTTC ACACAGATCT CACCCGTCTG GCTGCAGCTG AAGAGACGTG GCCGTGAGAT 660  
 GTTTGAGTTC ACCGGCTCTC ACAGCTGGA CCAAGGGTGG ATGCGAGCTG TCAGGAAGCA 720  
 60 TGCCAAGGTC CTGCACATAG TGCCTCGGCT CCGTGTGAG GACTGGACTT ACGATGATTT 780

300

CCGGAACGTC TTAGACAGTG AGGATGAGAT AGAGGAGCTG AGCAAGACCG TGGTCCAGGT 840  
 GGCAAAGAAC CAGCATTTCG ATGGCTTCGT GGTGGAGGTC TGGAAACCAGC TGCTAAGCCA 900  
 5 GAAGCGCGTG ACCGACCAGC TGGGCATGTT CACGCACAGG GAGTTTGAGC AGCTGGCCCC 960  
 CGTGCTGGAT GGTTCAGCC TCATGACCTA CGACTACTCT ACAGCGCATC AGCTGGCCCC 1020  
 10 TAATGCACCC CTGTCTGGG TTCGAGCGTG CGTCCAGGTC CTGGACCCGA AGTCCAAGTG 1080  
 GCGAAGCAAA ATCCTCCTGG GGCTCAACTT CTATGGTATG GACTACCGCA CCTCCAAGGA 1140  
 TGCCCGTGAG CCTGTGTGCG GGGCCAGGTA CATCCAGACA CTGAAGGACC ACAGGCCCCG 1200  
 15 GATGGTGTGG GACAGCCAGG YCTCAGAGCA CTTCTTCGAG TACAAGAAGA GCCGCAGTGG 1260  
 GAGGCACGTC GTCTTCTACC CAACCTGAA GTCCCTGCAG GTGCGGCTGG AGCTGGCCCC 1320  
 GGAGCTGGGC GTTGGGTCT CTATCTGGA GCTGGGCCAG GGCTGGACT ACTTCTACGA 1380  
 20 CCTGCTCTAG GTGGGCATTG CGGCCTCGC GGTGGACGTG TTCTTTTCTA AGCCATGGAG 1440  
 TGAGTGAGCA GGTGTGAAAT ACAGGCCTTC ACTCCGTAA AAAAAAAAAA AAAAAAAAAA 1500  
 25 AAAAAAAAAA AAAAA 1515

30 (2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 704 base pairs  
 (B) TYPE: nucleic acid  
 35 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

40 AAGATGGTGG CCCCCAGAGC TTGCTCTAT GCTGCTCCCC TGAGAGAGGC GTTTCATCA 60  
 ACCAGTTTTC CAAGGAGTTC AATGAGAGGA CAAAGGACAT CAAGGAAGGC ATTCTCTGC 120  
 CTACCAAGAT TTAGTGAAG CCTGACAGGA CATTGAAAT TAAGATTGGA CAGCCCACTG 180  
 45 TTCTCTACTT CCTGAAGGCA GCAGCTGGA TTGAAAAGG GCGCCGGCAA ACAGGGAAAG 240  
 AGGTGGCAGG CCTGGTGACC TTGAAGCATG TGTATGAGAT TGCCCGCATC AAAGCTCAGG 300  
 50 ATGAGGCATT TGCCCTGCAG GATGTACCCC TGTGCTCTGT TGTCCGCTCC ATCATCGGGT 360  
 CTGCCCCGTT TCTGGGCATT CGCGTGGTGA AGGACCTCAG TTCAGAAGAG CTTGCAGCTT 420  
 TCCAGAAGGA ACGAGCCATC TTCTGGCTG CTCAGAAGGA GGCAGATTTG GCTGCCCAAG 480  
 55 AAGAAGCTGC CAAGAAGTGA CCCTTGCCCC ACCAACTCCC AGATTTCAA GGAGGTAGTT 540  
 GCAAAAGCTG TGCCCAAGGG GAGGAAGGAG GTCACACCAA TATGATGATG GTTTTCATGA 600  
 60 CTTTGAATGA TATATTTTTC TACATCTAGC TGTATCGAGG CATCAGGCTT GAATTAACAT 660

CCTTTCTTAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA

704

5

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 1094 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

GGCAGCTTTC TTACAAACCC ATCCTTCTGA AATGTTGCTT CAAATTCATC CTCTGCTCCC	60
CAGTCCCACT ATTCCACACA TACTGTTACT GTTCTTTTAT CCTACTTTCT CAATTTTGGG	120
ACATAGTTGC AGTTACTGCA TTGAATACCT GTGGGTTTGC CTGTTGTTCT GTCTGTCTCT	180
GTGGTCTCTG TAATANTGGA TCCCAGAGAT AAAATGGACA GTTGTNATGC ACAGTTAATT	240
CAGAACTAG ACCTTACTTG CTGTGTGAAA TACCAACTAA ATTCTCAGTG AATCAGCTG	300
ANCTTATCT CCTTTTGTTC CCCCAATTG TATTTTCAGT TCAGGCCCCG AAAGATGGAA	360
TCCCAGCTAA GAANTACAAG TTACACCTG TACTAGCAGC CCATGTGTGC ATGTTCTTTA	420
AGTGTCTTTC CAGCTATGTC ATTTATATTC ATTTCCCTGT ATTATTATAA GCAAAGCATA	480
TTTGAGGAAA AAAACCCATA ATACCACACC TCATTTTTCG CAAGTAATAG GGTGATAAGT	540
CTCATCTCTC ATATAATATG TTGAGTATGC AGTATATTAT GTGTTAGGCT CTGGANAGGC	600
AGAGGTTAGA TCATGTWACA GATCATATCK GATTAGGCAG ATAAACAGTA TTTTAACCTT	660
TTCTTTATTA TATGTAACTT GCTTTCAGGT TTTTAAATGT TACTATTATG TCTTTAATAT	720
ATTATCTTTA TTTGTACTTT TGTATACAGA GTGATTTTCC TTTTAAAAA AAAATTGTGT	780
CTTTAGGATG GATTCCAAAG ATGTGGAATC AGTAGGTTTA AGGAATATGG ATATTTTGGC	840
TGGCAAGGTG GCTCACACCT GTAATCCGAG CACTTTGGGA GGCTGAGGTG GGTGGATCAC	900
CTGAAGTCAG GAGTTCGAGA CCAGCCTGAC CAACATGGCG AAACCTGTT TTTACTAAAG	960
ACACACWAA AATTGGCCAG TGGTGGTGGC ATGTGCTTGT AGTCCCACTT AGCTACTCGA	1020
GAGGCTGAGG CAGGAGAATC GCTTGAACCC GGGAGGCAGA GGTGCACTG AGGCAAGATG	1080
GCACCTCTAC ACTC	1094

55

(2) INFORMATION FOR SEQ ID NO: 43:

60

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1321 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(E) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

	TGGCTTAGGC CAGCAGCTTC GCGTTGGCTG GAACTACTGG ACAGACCCTT TTGAGATGTG	60
10	CCTGTGGTGC TGTGAGATG TGTGTAGTGG TCTTAGCTCT TTGTTGAGCT TGTGTGTGTG	120
	TTGTGTAGTC TTATGCTGAT GGTGAATATG GCGGTGTGTT GGAGGGCTTC TTAGCTCTTT	180
15	GGTGAGATTC TATTTCTATG TGTTTGTATC ASCTGAATGT TGCTGGAAT AAAACCTTGG	240
	TTTGTGAAGG CTCATTTTTC TGGGAGTAA GTACGGGAAA AGGTCTTTGA GGGTTCCTAG	300
	GCTGCTTTGT ACAACAGGAA AATGCTCAA AGCCTTGCTT CCCAGCAACT TGGGGCTGGT	360
20	TCCCAGTACC TGGTCTGACC GCTTCCTGGT TCTTATCTCA AGGCAGAGCT TCTGAATTTT	420
	AGGCCTTCAT TCCAGAGCCC TCTTGTGGCC AGGCCTTCCT TTGCTGGAGG AAGGTACACA	480
25	GGTGAACTT GATGCTGTAC TTGGGGGATC TCCTTGGCCT GTTCCACCAA GTGAGAGAAG	540
	GTACTTACTC TTGTACCTCC TGTTCAGCCA GGTGCATTAA CAGACCTCCC TACAGCTGTA	600
	GGAATACTG TCCAGAGCT GAGGCAAGGG GATTTCTCAG GTCATTTGGA GAACAGTGC	660
30	TTTATAGTA GTTAAAGTA GTAATGCTA CTGTATTTAG TGGCGTGGAA TTCAGAAGAA	720
	ATTGTAAGAC CAGAGCAGGG GTGGTCTGCA TGTGAATGAA CAGGAATGAG CCGGACAGGC	780
35	TGGCTGTGAT TCGTTTCTTC CTCGCCATTT GGACCTTCT CTGCCCTTAC ATTTTGTGTT	840
	CTGCATCTAC CAGCATCCAC CAGTCTATTT ATTAAGTTAG CAAGAGGACA AGTAAAGGGC	900
	CCTCTTGGCT TGAATTTGCT TCTTTCTTTC TGTGGAGGAT ATACTAAGTG CGACTTTGCC	960
40	CTATCCTATT TGGAAATCCC TAACGAAATT GAGTTTCTTA TTAAGGATCC AAAAAGAAAA	1020
	ACAAAATGCT AATGAAGCCA TCAGTCAAGG GTCACATGCC AATAAACAAT AAATTTTCCA	1080
45	GAAGAAATCA AATCCAATA GACAAATAAA GTAGAGCTTA TGAAATGGTT CAGTAAGGAT	1140
	GAGTTTGTG TTTTGTGTTT TGTTTGTGTT TGATTTTTTA AAGACGGAGT CTCGCTCTGT	1200
	CAGTCAGCT GGAATGCACT GGTATCATCT TGGCTCACTG TAACCTCCGC CTCCCGGGTT	1260
50	CAAGCCATTC TCGTCCCTCA GTCTCTGAG TAGCTGGGAT TACAGGTGCG TGCCACCATG	1320
	CCTGGCTAAT TTTGTGTTT TTAGTAGAGA CAGGGTTTCA CCAATGTTGT CGGGCTGGTC	1380
55	TCAAACTGCT GACCTCTTGA TCGGCTGGCC TTGGCTCCC AAAGTGATGG GATTACAGAT	1440
	GTGAGCCACC CGTCCCTAG CCAAGGATGA GATTTTAAA GTATGTTTCA GTTCTGTGTC	1500
	ATGGTTGAAA GACAGAGTAG GAAGGATATG GAAAAGGTCA TGGGAAGCA GAGGTGATTC	1560
60	ATGGCTTGT GAAATGAGG TGAATGTTT CTTATGCTCT AGGCCACTTG TGAAGAATAT	1620

303

GAGTCAGTTA TTGCCAGCCT TGGAAATTTAC TTCTCTAGCT TACAATGGAC CTTTTGAACT 1680  
 5 GGAAAACACC TTGTCTGCAT TCACCTTAAA ATGTCAAAAC TAATTTTAT AATAAATGTT 1740  
 TATTTTCACA TTGAAAAAA AAAAAATTT AAAAACYCGG GGGGGGGGCS G/AACCCCAT 1800  
 NGCCCTAAG GGGGGGGTT T 1821

## (2) INFORMATION FOR SEQ ID NO: 44:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1024 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GGGGCACAGT TGAAGAAGCG ACCGAGGGAC TGGGAGTCGT TAGTGAGGAT GACGGCGCAT 60  
 25 GGCAAGAACT GCACCGCAGG GCGTCTACA CCTACCCAGA GAAGAAGAG GACACAGCGG 120  
 CCTCGGGCTA TGGGACCCAG AACATTCGAC TGAGCCCGGA TCCCGTGAAG GACTTCGACT 180  
 30 GCTGTGTGCT CTCCCTGCAG CCTTCCCACG ATCTGTGTGT CACCCAGAT GGCTACCTGT 240  
 ATGAGCGTGA GGCATCTTG GAGTACATTC TGCACCAGAA GAAGGAGATT GCCCGGCAGA 300  
 TGAAGGCTA CGAGAAGCAG CGGGGCACCC GCGCGGAGGA GCAGAAGGAG CTTGAGCGGG 360  
 35 CGGCCTCGCA GGACCATGTG CGGGCTTCC TGGAGAAGGA GTCCGCTATC GTGAGCCGGC 420  
 CCCTCAACCC TTTCACAGCC AAGGCCCTCT CGGGCACCAG CCCAGATGAT GTCCACCTG 480  
 40 GGCCAGTGT GGTCTCTCCA AGTAAGGACA AGGACAAAGT GCTGCCAGC TTCTGGATCC 540  
 CGTCGCTGAC GCCCGAAGCC AAGGCCACCA AGCTGGAGAA GCGTCCCCC ACGGTGACCT 600  
 GCCCCATGTC ACGGAAGCCC CTGCGCATGT CGGACCTGAC GCGCGTGCAC TTCACACCGC 660  
 45 TAGACAGTTC CGTGACCGC GTGGGGCTCA TCACCCGCAG CGAGCGCTAC GTGTGTGCGG 720  
 TGACCCCGGA CAGCCTGAGC AACGCCACCC CCTGCGCTGT GCTGCGGGCC TCTGGGCTG 780  
 50 TGGTCACCTT CGAATGCGTG GAGAAGCTGA TTCGGAAGGA CATGGTGGAC CCTGTGACTG 840  
 GAGACAACT CACAGACCGC GACATCATCG TGCTGCAGCG GGGCGGTACC GSTTCGCGGG 900  
 CTCGGGAGTG AAGCTGCAAG CGGAGAAATC ACGGCCGGTG ATGCAGGCCT GAGTGTGTGC 960  
 55 GGGAGACCAA ATAACCGGC TTGGGTGCGC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1020  
 AAAA 1024

60

## (2) INFORMATION FOR SEQ ID NO: 45:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 983 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

CGACACGGCT GCGAGAAGAC GACAGAAGGG CCCGACCGCG AGCCGTCCAG GTCTCAGTGC 60  
 TGTGCCCCCC CCAGAGCCTA GAGGATGTTT CATGGGATCC CAGCCACGCC GGGCATAGGA 120  
 GCCCCCTGGGA ACAAGCCGGA GCTGTATGAG GAAAGTGAAGT TGTACAAGAA CCCCCGGGAG 180  
 AGGGAGAAGT ACGACAACAT GGCAGAGCTG TTTGCGGTGG TGAAGACAAT GCAAGCCCTG 240  
 GAGAAGGCCT ACATCAAGGA CTGTGTCTCC CCCAGCGAGT ACACTGCAGC CTGCTCCCGG 300  
 CTCTCGTCC AATACAAAGC TGCCTTCAGG CAGGTCCAGG GCTCAGAAAT CAGCTCTATT 360  
 GACGAATTCT GCCGCAAGTT CCGCCTGGAC TGCCCCGTGG CCATGGAGCG GATCAAGGAG 420  
 GACCGGCCCCA TCACCATCAA GGACGACAAG GGCAACCTCA ACCGCTGCAT CGCAGACGTG 480  
 GTCTCGCTCT TCATCACGGT CATGGACAAG CTGCGCCTGG AGATCCGCGC CATGGATGAG 540  
 ATCCAGCCCCG ACCTGCGAGA GCTGATGGAG ACCATGCACC GCATGAGCCA CCTCCCACCC 600  
 GACTTTGAGG GCCGCCAGAC GGTGAGCCAG TGGCTGCAGA CCCTGAGCGG CATGTCCGGC 660  
 TCAGATGAGC TGGACGACTC ACAGGTGCGT CAGATGCTGT TCGACCTGGA GTCAGCCTAC 720  
 AACGCCTTCA ACCGCTTCCT GCATGCCTGA GCCCGGGGCA CTAGCCCTTG CACAGAAGGG 780  
 CAGAGTCTGA GCGATGGCT CCTGTCTCCC TGTCGGCCAC ACAGGCCGTG GTCATCCACA 840  
 CAACTCACTG TCTGCAGCTG CTTGTCTGGT GTCTGTCTTT GGTGTCAGAA CTTTGGGCC 900  
 GGGCCCCCTCC CCACATAAA GATGCTCTCC GACCTTCAA AAAAAAAAAA AAAAAAAGR 960  
 KGSGGCCGGT CCCCANTCCC CCC 983

## (2) INFORMATION FOR SEQ ID NO: 46:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2421 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

CCGGCTGATC GCTGCGGCTC CGCCAATACA ATAGAGCCAK CCACTACCAG CAGCCTGGCC 60



	CTCTTCCTCC TTCTCCAGAG AGACCAATCC AGCCGAACTC GGGGTTTGCC TGAGGAGAAG	120
	GAGGAAGTGA CCATGGACAC AAGTGAAAAC AGACCTGAAA ATGATGTTCC AGAACCTCCC	180
5	ATGCCTATTG CAGACCAAGT CAGCAATGAT GACCGCCCGG AGGGCAGTGT TGAAGATGAG	240
	GAGAAGAAAG AGAGCTCGCT GCGCAATCA TTCAAGAGGA AGATCTCCGT TGTCTCAGCT	300
10	ACCAAGGGGG TGCCAGCTGG AAACAGTGAC ACAGAGGGGG GCCAGCCTGG TCGGAAACGA	360
	CGCTGGGGAG CCAGCACAGC CACCACACAG AAGAAACCTT CCATCAGTAT CACCACTGAA	420
	TCACTAAAGA GCCTCATCCC CGACATCAAA CCCCTGGCGG GGCAGGAGGC TGTGTGGAT	480
15	CTTCATGCTG ATGACTCTCG CATCTCTGAG GATGAGACAG AGCGTAATGG CGATGATGGG	540
	ACCCATGACA AGGGGCTGAA AATATGCCGG ACAGTCACTC AGGTAGTACC TGCAGAGGGC	600
20	CAGGAGAATG GGCAGAGGGA AGAAGAGGAA GAAGAGAAGG AACCTGAAGC AGAACCTCCT	660
	GTACCTCCCC AGGTGTCACT AGAGGTGGCC TTGCCCCCAG CTGCAGAGCA TGAAGTAAAG	720
	AAAGTGACTT TAGGAGATAC CTTAACTCGA CGTTCCATTA GCCAGCAGAA GTCCGGAGTT	780
25	TCCATTACCA TTGATGACCC AGTCCGAACT GCCCAGGTGC CCTCCCCACC CCGGGGCAAG	840
	ATTAGCAACA TTGTCCATAT CTCCAATTG GTCCGTCTTT TCACTTTAGG CCAGCTAAAG	900
30	GAGTTGTGGG GCGGCACAGG AACCTTGGTG GAAGAGGCCT TCTGGATTGA CAAGATCAAA	960
	TCTCATTGCT TTGTAACGTA CTCAACAGTA GAGGAAGCTG TTGCCACCCG CACAGCTCTG	1020
	CACGGGGTCA AATGGCCCCA GTCCAATCCC AAATTCCTTT GTGCTGACTA TGCCGAGCAA	1080
35	GATGAGCTGG ATTATCACCG AGGCCTCTTG GTGGACCGTC CCTCTGAAAC TAAGACAGAG	1140
	GAGCAGGGAA TACCACGGCC CCGGCACCCC CCACCCCCAC CCGCGGTCCA GCCACCACAG	1200
40	CACCCCCGGG CAGAGCAGCG GGAGCAGGAA CGGGCAGTGC GGAACAGTG GGCAGAACGG	1260
	GAACGGGAAA TGGAGCGGCG GGAGCGGACT CGATCAGAGC GTGAATGGGA TCGGGACAAA	1320
	GTTCGAGAAG GSCCCGTTT CCGATCAAGG TCCCGTRACC GCGCCCGCAA GGAACGTCCG	1380
45	AAGTCTAAAG AAAAGAAGAG TGAGAAGAAA GAGAAAGCCC AGGAGGAACC ACCTGCCAAG	1440
	CTGCTGGATG ACCTTTTCCG AAAGACCAAG GCAGCTCCCT GCATCTATTG GCTCCCACTG	1500
50	ACTGACAGCC AGATCGTTCA GAAAGAGGCA GAGCGGGCCG AACGGGCCAA GGAGCGGGAG	1560
	AAGCGGCGAA AGGAGCAAGA AGAPGAAGAG CAAAAGGAGC GGGAGAAGCA AGCCGAGCGG	1620
	GAACCGAACC GACAGCTGGA GCGAGAGAAA CGTCGGGAGC ACAGTCGGGA GAGGGACAGG	1680
55	GAGAGAGAGA GAGAAAGGGA GCGGGACAGG GGGGACCGAG ATCGGGATAG GGAAGGGAC	1740
	CGAGAACCAG GCAGGGAAAG GGATCCGAGG GACACCAAGC GCCACAGCAG AAGCCGGAGT	1800
60	CGGAGCACAC CTGTGCGGGA CCGGGGTGGG CGCCGCTAGC TGGGAAAACA CTAGAGCTGC	1860

AGGTACCAGC CACTCGGCCC CAGGGGGTTA TGGCCACAGA GGGATAGGCA CAGTCTCCAC 1920  
 CACCCTGGAG CCAAGGGTCT TTCACATCAC CTATCCCTAC ATACATACCA AATGGAAAAG 1930  
 5 TGGCCATCCT TTTCCCCCA AACACACCCC CTAACTCTAT CTCTTGGGAC TTAGCCCGAC 2040  
 CCTCCCTCTC ATTCCCAAT AAGTCTGAGA GGCAAGAGCT AGGTACGCA AGGAGGTGGT 2100  
 TGGCCAGAGA TGGGGAACAG CCAGGTGCCC CAGTCTCTG ATTTTCTCTC CATCCTGCTT 2160  
 10 ACCACCTCCC TGGTACTTA CAGCCTTCTC TTGGGAACAG CCGGGGCCAG GACTGGGTCA 2220  
 CCTATGAGCT GAATCAGCAT CTCTCTCTGA GTCCACGGCC CCTGCAGTT CCCAGTCTCT 2280  
 15 TCTGTCTTGC AGCCCTTGCC TCTTTCCAC AGGTTCCACT TTATATCCAC CTTTTCCTTT 2340  
 TGTTCATTTT TTATTTTAT TTTTATTAT ATTAAATGAT GTGGTCTATG GAAAAAAAAA 2400  
 TAAAAATCTG ACTTAGTTTT A 2421  
 20

## (2) INFORMATION FOR SEQ ID NO: 47:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 840 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

CTCAAACTCC TGACCTGAAG CGATCTACCT GCCTCAGCTA GGATTACAGG TGTGAGCCAC 60  
 35 CGCACCCAAC CTCAATAAGC KTATTTGATA AAKATATGC AAGTCCCTT TATKCACTTT 120  
 TCATTCAGAA TGTTAGTAA TTTGTATTGT TTTTCAGATT TTCAGCCCA TATATCTCTT 130  
 40 TGGCCACTGT GTCACTGTAT TCTACCTAWA CATCATCAG TGTTCCTGCT ATTGGCTGTA 240  
 TGATGGAACA CTGGGGCTCA TTTTCTGAA AACTGCCGAT AGTGCATAGA RTGCTGGGAT 300  
 GGAAACCAGA ARCTTTGAAT TCAAGCCTTG GTTCTGCCTT GTTTTTCCTT GGGTGGCCTT 360  
 45 GAGTCAGCCA CATACCTTTT AAAATCTCAA TTTATTAGAA ATTATTCCAA ATCAAAATCA 420  
 AATGAGAAGG TATATACAAA AGTGCTTTAT CCCACAATAA ACTATTCAAG AGAGAGCAAA 480  
 50 GGAGAGGACA TTTACTCAAC ACCTCCTAAA AGGCAGCCAG TGAAATTAGG CATTTTATTT 540  
 AATCCTCCTG GCAACTCTGA GAGTAAAGCA TTATTAATCC CATTTGGCT GTTTAAAGCA 600  
 ATTATTTGCA CTAGATTCCA GCTGTAGTTT AGTTTCAGAA AAAAAATCC TGAGATGTGA 560  
 55 ATTCACAGCT TTCTGGGTTT AAAGCCCAAG CTCTATCACA TCATGCTATT ATTGTTACAT 720  
 TACTGCTAGT TCTATGAAAA GAATACTAA TTTATGAAAT ACATCTTATC CAAAAAAAAA 780  
 60 AAAAAAAAC TGGGAGGGGG GGGCCGTACC CAAATCGCCG GATAGTGATC GTAAACAATC 840

## 5 (2) INFORMATION FOR SEQ ID NO: 48:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 2432 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

15 GGCACGAGGC CCGAACGCT GAGGAAGGGC CCGTCCCGCC TTCCCCGGCG CGCCATGGAG 60  
CCCCGGGGCG TTGCAGAAGC CGTGGAGAGC GGTGAGGAGG ATGTGATTAT GGAAGCTCTG 120  
CGGTCAATACA ACCAGGAGCA CTCCAGAGC TTCACGTTTG ATGATGCCCC ACAGGAGGAC 180  
20 CGGAAGAGAC TGGCGGASTG CTGGTCTCCG TCCTGGAACA GGGCTTGCCA CCTCCACACC 240  
GTGTCACTTG GCTGCAGAGT GTCCGAATCC TGTCCCGGGA CCGCAACTGC CTGGACCCGT 300  
25 TCACCAGCCC CCAGAGCCTG CAGGCAYTAG CCTGYTATGY TGACATCTCT GTCTCTGAGG 360  
GGTCCGTCCT AGAGTCCGCA GACATGGATG TTGTAAGTGA GTCCCTCAAG TGCCTGTGCA 420  
ACCTCGTGCT CAGCAGCCCT GTGGCAGAGA TGCTGGCAGC AGAGGCCCCG CTAGTGGTGA 480  
30 AGCTCACAGA GCGTGTGGGG CTGTACCGTG AGAGGAGCTT CCCCCACGAT GTCCAGTTCT 540  
TTGACTTGCG GCTCCTCTTC CTGCTAACGG CACTCCGCAC CGATGTGGCG CANAGCTGTT 600  
35 TCAGGAGCTG AAAGGAGTGC GCCTGCTAAC TGACACACTG GAGCTGACCG TGGGGGTGAC 660  
TCCTGAAGGG AACCCCCCAC CCAGGCTCCT TCCTTCCCAA GAGACTGAGC GGGCCATGGA 720  
GATCCTCAAA GTGCTCTTCA ACATCACCTT GGAATCCATC AAGGGGGAGG TGGACGAGGA 780  
40 AGACGCTGCC CTTTACCGAC ACCTGGGGAC CTTTCTCCGG CACTGTGTGA TGATCGCTAC 840  
TGCTGGAGAC CGCACAGAGG AGTTCCACGG CCACGCAGTA ASCCTCCTGG GGAACCTTGGC 900  
45 CCTCAAGTGT CTGGATGTTT TCCTCACCTT GGAGCCACAT GGAGACTCCA CGGAGTTTAT 960  
GGGAGTGAAT ATGGATGTGA TTCGTGCCCT CTCATCTTC CTAGAGAAGC GTTGTGACAA 1020  
GACACACAGG CTGAAGGAGA GTGTAGCTCC CGTCTGAGC GTGCTGACTG AATGTGCCCC 1080  
50 GATGCACCGC CCAGCCAGGA AGTTCTTGAA GGCCAGGTG CTGCCCCCTC TGGGGGATGT 1140  
GAGGACACGG CCTGAGGTTG GGGAGATGCT GCGGAACAAG CTTGTCCGCC TCATGACACA 1200  
55 CCTGGACACA GATGTGAAGA GGGTGGCTGC CGAGTTCTTG TTTGTCTGT GCTCTGAGAG 1260  
TGTGCCCCGA TTCATCAAGT ACACAGGCTA TGGGAATGCT GCTGGCCTTC TGGCTGCCAG 1320  
GGGCTCATG GCAGGAGGCG GCGCGAGGCC AGTACTCAGA GGATGAGGAC ACAGACACAG 1380  
60

ATGAGTACAA GGAAGCCAAA GCCAGCATAA ACCCTGTGAC CGGGAGGGTG GAGGAGAAGC 1440  
 CGCCTAACCC TATGGAGGGC ATGACAGAGG AGCAGAAGGA GCACGAGGCC ATGAAGCTGG 1500  
 5 TGACCATGTT TGACAAGCTC TCCAGGAACA GAGTCATCCA GCCAATGGGG ATGAGTCCCC 1560  
 GGGGTGATCT TACGTCCCTG CAGGATGCCA TGTGCGAGAC TATGGAGCAG CAGCTCTCCT 1620  
 CGGACCCTGA CTCGGACCCT GACTGAGGAT GGCAGCTCTT CTGCTCCCCC ATCAGGACTG 1680  
 10 GTGCTGCTTC CAGAGACTTC CTGCGGGTTC CAACCTGGGG AAGCCACATC CCACTGGATC 1740  
 CACACCGGCC CCCACTTCTC GATCTTAGAA ACCCCTTCTC TTGACTCCCG TTCTGTTTAT 1800  
 15 GATTTGCCTC TGGTCCAGTT TCTCATCTCT GGAAGTCAAC GGTCTTCTTG TGCTAGAAGT 1860  
 CAGGCTCAGC CTCGAATTCC ACAGACGAAG TACTTTCTTT TGTCTCGGCC AAGAGGAATG 1920  
 TGTTTCAAG CTGCTGCCTG AGGGCAGGGC CTACCTGGGC ACACAGAAGA GCATATGGGA 1980  
 20 GGGCAGGGGT TTGGGTGTGG GTGCACACAA AGCAAGCACC ATCTGGGATT GGCAGACTGG 2040  
 CAGAGCMANT GTKTTGGGGT ATGTGCTGCA CTTCCAGGG AGAAAACCTG TCAGAAGTTT 2100  
 25 CCATACGAGT ATATCAGAAC ACACCCCTCC AAGGTATGTA TGCTCTGTTG TTCCTGTCTT 2160  
 GTCTTCACTG AGCGCAGGGC TGGAGGCCTC TTAGACATTC TCCTTGGTCC TCGTTCAGCT 2220  
 GCCCAGTGTG GTATCCACAG TGCCCGAGTT CTCGCTGGTT TTGGCAATTA AACCTCCTTC 2280  
 30 CTACTGGTTT AGACTACACT TACAACAAGG AAAATGCCCC TCGTGTGACC ATAGATTGAG 2340  
 ATTTATACCA CATACCACAC ATAGCCACAG AAACATCATC TTGAAATATA GAAGAGTTTT 2400  
 35 GGACAAAAAA AAAAAAAAAA AAAAAAAAAA AA 2432

40 (2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1742 base pairs  
 (B) TYPE: nucleic acid  
 45 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

50 GTCTGTCAGG AGCTGCACGC GGGCGAGGTG CGCANGAACA AGGAGCAGCG AGAAGAGATG 60  
 TCGGGCTAAG GGGCCGSAC GGGGGGGCC CATCTGCGA CGGAACACGT TCGGGTTTTG 120  
 GTTTTGTTC GTTCACTCT GTCTAGATGC AACTTTTGTT CTCCTCCCC CACCCAGCC 180  
 55 CCCAGCTTCA TGCTTCTCTT CCGCACTCAG CCGCCCTGCC CTGTCTCTGT GGTGAGTCGC 240  
 TGACCACGGC TTCCCTGCA GGAGCGCGG GCGGTGRAGA CGCGTCCCT CGGTGCAGAC 300  
 60 ACCAGGCGCG GCGCGGTGG GTCCCCGGG GGCCTGTGA GAGAGGTGG GGTGACCTG 360

GTAAACCCAG GCGCGTGGCG TGGGATCRGG GGTCCCTTACG CTGGGCTGTC TGGTCAGCAC 420  
 GTGCAGGTCA GGGCAGGTCC TCTGAGCCGG CGCCCTGGC CAGCAGGCGA GGCTACAGTA 480  
 5 CCTGCTGTCT TTCCAGGGGG AAGGGGCTCC CCATGAGGRA GGGGCGACGG GCGAGGGGGG 540  
 TGATGGTGCC TGGGAAGCCT GCKTGTGCAN CCGGTGCTTG TTGAACTGGC AGGCGGGTGG 600  
 10 GTGGGGGCTG CAGCTTTCTT TAATGTGGTT GCACAGGGGT CCTCTRAGAC CACCTGGCGT 660  
 GAGGTGGACA CCCTGGGCCT TCCTGGAAGC CTGCAGTTGG GGGCCTGCCC TGAGTCTGCT 720  
 GGGGAGTGGG CATCTCTGTC CAGGGACCCA TGAGCAGGCT GCATGGTCTA GAGGTTGTGG 780  
 15 GCAGCATGGA CAGTCCCCCA CTCAGAACTG CAAGAGTTCC AAAGAGCCTC TGGCCCAGGC 840  
 CCCTCGGTGG GACAGCCCCG CCGCCCTCC CCACCAGGGC TTTGCAGATG TCCTTGAAAG 900  
 20 ACCCACCCTA GAGCCCTTTG GAGTCTGGC CCCTCTGTG CCCTCTGCCC TGGTGAAGC 960  
 GGCASCACAA GTCTCTCTCA GGGAGCCCCA AGGGGGATT TKTGGGACCG CTGCCCACAG 1020  
 ATCCAGGTGT TGAAGGGCA GCGGGTAAGG TTCCAAGCC AGCCCCAACA CCTTCCCAC 1080  
 25 TTGGCACCCA GAGGGGGCTG TGGGTGGAGG CTTGACTCCA GGCTCTCCT GCCCACCCC 1140  
 TCTGGGCTGA GTTCTTCTT TCCCTTGGAC GCCCAGTGT GGCCTTGGAG GACGGTLAGC 1200  
 30 TGGAGGATGG CCGTGGGGGA GGCTGTCTTT GTACCACTGC AGCATCCCC ACTTCTCCAC 1260  
 GGAAGCCCCA TCCCAAAGCT GCTGCCTGGC CCCTTGTGT AAAGTGTGAA GGGGGCGGCT 1320  
 GAGTTCTCTT AGGACCCAGA GCCAGGGCCC TCAACTTCCA TCCTGCGGGA GGCCTTGGCC 1380  
 35 GGGCACTGCC AGTGTCTTCC AGAGCCACAC CCAGGGACCA CGGGAGGATC CTGACCCCTG 1440  
 CAGGGCTCAG GGGTCAGCAG GGACCCACTG CCCCATCTCC CTCTCCCCAC CAAGACAGCC 1500  
 40 CCAGAAGGAG CAGCCAGCTG GGATGGGAAC CCAAGGCTGT CCACATCTCG CTTTGTGGG 1560  
 ACTCAGAAAG GGAAGCAGAA CTGAGGGCTG GGATATTCTT CATGGTGGCA GCGTTCATAG 1620  
 CGAAAGCCTA CTGTAATATG CACCCATCTC ATCCACGTAG TAAAGTGAAC TTAAAAATTC 1680  
 45 AATCAAATGA ACAATTAAAT AAACACCTGT GTGTTTAAGA AAAAAAAAAA AAAAAAACTG 1740  
 CG 1742

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1487 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

	GGCACGAGCC TCCGCGAACT GTGGAGTCGG CGGAGGGCTG GAATCAGCGT GGGCTCCAGG	60
5	TGGCTGSCAG CCGGGTGGCA GAACTCTTCC GAGGCTCCTT GGAAGAAGC TACACCCGAG	120
	GGAGCCCGAT GGGCCTCGAA AACCTGGCCC GCTCTGGTTC TGTACCATTG CAAGGGGAAC	180
10	CGTAAACTGA GCTTTTCTAA CGTGGGTTC TGCCAAGTAC TTTCCAGCT GCCCCTTCC	240
	CCCCAGCACA CAGGAGAGCC TCTGTGTAGC CAGCGCTTGA CAGTCGTTAG GTAGGTTGTA	300
	CTGTGTAGGG AGGAGCTCAA GATCATGAAT GGTGTGACA GGAGAAAGCG GTTGCATCTT	360
15	TGCAAACTA TATACCTGCT GTGGTTTGTG TTTCTTTTC TGCTGAGTAA TGAAGTTGTA	420
	AGTTCACACT GGCACATTCT CAGGGCTGTG CAGATTATTT GCACTTTATT TCATAGGTGR	480
20	ATAAGTGCTT TTTAGCTTTC TTTGTATATT GAGTGTCTTT TGAATTGCTT CCCATATTTT	540
	TATTTCATAC AAAGTGAACA ATTGTGGCCC CTCTATTTTA TTTATAAAGG TTCAGTGTAT	600
	CTTTGCCTGC CTACATCAAT CTGCAAGGGA GTTGCAGAAA GCCTCATGTT CATCGAGCCG	660
25	TGAGTCACAA CCAATTCTA AGCTGTTATA ACAAAAAAGT GTTGTCTTTT TTTCAAGT	720
	AACCTTAAAA GTGTAGTTTA GAAAGAAAC ATTTTCAATA AAAAGACACT ACATTAATCC	780
30	TGGATGCTTG CAAATCCTAA AATMTATTCC TCCTCTAGCG TTGCACAGCT CTGTGTGTA	840
	TACACAGACT AGCTTTAAAA TTTGTACAT ACCACTTTAC CTTTACTTTT ATGTATCATT	900
	CCCCCGACTT CCTTACTGCA GGTGTGGGCA AGAAAACTTT TCCTTTAACA CTTTCAACA	960
35	GCGGGCATAA AATTCTGCAG CTGAGGTCTT GAAGAATGCA GATGGGTACA GTATGTGTTG	1020
	GAGCTCACAG TGTGTATTGA CTAACCTAGT TCCTTTTTTG CTTTTTTTGG TATTGTCTTG	1080
40	TTAAAAGTGA CTCCCAGGTA GCAACTCTCT TTTTAAGGG TGGGAACGAA AGGGACGTAG	1140
	GAAGAATAGA TCTAGATTAT TTAACAGTCT TCGATAGAGT TTGAAAGCTT TCTTCTTCAT	1200
	TCAATTTTGG GCAAAATACT GCCTCTGCAT TTGTTCATAA CAAAAAGATT AGATTAATAA	1260
45	GTAGCTTTTG TTGGTGGAAA TTACCAGCTC TATAAGTCAC CCTTGGTGGT TCATGGACCT	1320
	CTGATTAGCT TGGGTTTTGC AGTCTCATTG CCACATGTAT ATGTGGAGCC AATGGCCTTT	1380
50	TGGTGCTCAG CTGTTTACGT CTGACTCCTT GACTTCTTTG GTACAGTGAT GGAGTCAGAT	1440
	CTCATTAAGT GTGATTCTCC ATGGATATAA CCAGCCCCAA AAAAANG	1487

55

## (2) INFORMATION FOR SEQ ID NO: 51:

## (i) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 1328 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

5  
GGCAGGAGCT CGTGGCGAAT TCGGCACGAG AGAAGATTTC AAGAGCCAG ATCCAGCTTC 60  
CCTGCGGGCT GCTTCTTGTC GGAAGGGAA AAAGAGGAAG GCCTGTAAGA ACTGCACCTG 120  
10 TGGCCTTGCC GAAGAAGTGG AAAAAGAGAA GTCAGGGAA CAGATGAGCT CCCAACCCAA 180  
GTCAGCTTGT GGAACTGCT ACCTGGGCGA TGCCTTCGGC TGTGCCAGCT GCGCCTACCT 240  
TGGGATGCCA GCCTTCAAAC CTGGGAAAAA GGTGCTTCTG AGTGATAGCA ATCTTCATGA 300  
15 TGCCTAGGAG GTTCTGACA TGGGACCCAT CTGCTCCTCC AGCCAACTCC TGTCCCTCAC 360  
ATCCCACCAT GGTGGCTCCT CCCACCTCCT CTGGATTGT TCACTCTGAG ATCTGTTTGC 420  
20 AGAGTGGGTG CTTAGCAGAC AGAGTGAAGC TGGCTGGGGG GCACAGTGGT GTGTAGTGCT 480  
GCTGTGTATC AAAAGACCAA GGTATTATGG GACCTGGTTT CAGAATGGGA TGGGTTTCTT 540  
CACCTCATGT TAAGAGAAGG GAGTGTGTCC TGAAGAAGCC CTTCTTCTGA TGTAAAATG 600  
25 CTGACCAGAA CGCTCTTGAG CCCAGGCATC GTTGAGCATT AACACTCTGT GACAGAGCTG 660  
CAGACCCCTG CCTTGAGTCT CATCTCAGCA ATGCTGCCAC CCTCTTGTCT TTCAGAGTTG 720  
30 TTAGTTTACT CCATCTTTG TGACACGAGT CAAGTGGCTC ACAACCTCCT CAGGGCACCA 780  
GAGGACTCAC TCACTGGTTG CTGTGATGAT ATCCAGTGTC CCTCTGCCCC CTTCATCCC 840  
CAACCACATT TGAAGTAGC ATTGCATCTG TGTCTGTTG TCATTTATGT TAACCTTCAG 900  
35 GTATTAACT TGCTGCATAT CTTGACATAT CTTGAGATTC TGCATGTCTT GTAAAGAGAG 960  
GGGATGTGCA TTTGTGTGTG ATGTTGGATA GTCATCCAGC CTCAGTTTGG ACCATTGGAG 1020  
40 GAAGTTAGTG TCACGCACAA ATGGGGCTAT TCCTACGCTT AGAATAGGGC TTGTCTGCCC 1080  
ACTTTAGAAG AGTCCCAGGT TGGTGAGCAT TTAGAGGGAA GCAGGGCAGA ACTCTGAACG 1140  
ACAATACGTC TCTCTGAGCA GAGACCCCTT TGTCTTGTAT ATCCACCCAT ATGGACTTGG 1200  
45 AATCAATCTT GCCAAATATT TGGAGAGATT GTGTGGATTT AAGAGACCTG GATTTTATA 1260  
TTTTACCACT AAATAAAAGT TTTCATTGAT ATCTGTCCTT GAAAAAAAAA AAAAAAAAAA 1320  
50 AAACGCA 1328

55 (2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1856 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

## (D) TOPOLOGY: linear

(M1) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

5	GAATTGGGCA CGAGCTGTGC AGCATTTGGA ATGACTTGGC ATTCGAGGGT TCGGCTGGCC	60
	CCTAGATTAA ATTCCCGGGG CTGAAATTTG GTTGCAGATT TATATATCA TATTTTAAT	120
	TGCTGTCTTC AATTAAAGCA TTATGAGCA TAAATATTT TCGATGTGC GATGCATGCT	180
10	TTTCAGGGCC TTCTTTCTTT GTACAAAT ATATGTGCA AAGCTTTTC ACTTATATTC	240
	TTCAACATG ATGTAATTT AATTAATTA CTCTATGA TATTTATTA TTCCTATGAT	300
15	TTTGGCACTG TTATATCTTC TCTCAAAAT ACATCTGGG AAGAGATTA TTTAAGTGA	360
	TTTGATTATC TTCTATCTTC TTTATTTAT TTCTATTA CTTAAGAAAT TCGTTCCATT	420
	GGTTGGCATT GATACAGTAA ATTTGTAAAT GAGGAGACA TAAAAAAT CTAAATTACT	480
20	TGTGCTTAAT GACTGTAGCA GAATSCCTTT TCTTAATTC AGATGTCTT TCTTGCAGTT	540
	TAGTTTGATA GATTGTCAAG CTATGCTGTT TCGATGAGT TATTTGGCT GGTAGGAACG	600
25	CAGGCTTTCT TGTCTCTGGT TGTAGCTTGC ATATCTGCT CATTAGGCA ACAACGTAGC	660
	CGGAGATCAC AATCAGGGC CTTCGTTTAT TTCTATTTT CTGAGGTGC AGAGAGGTTG	720
	GCAAGAACTG ACCTCACTGG GCAAGGTTGG CCACTGACT GATTTTTTA TGCACCTAT	780
30	GTGTTCAAGG AGCCACAGGC CATATTTTAC TCTCAAAAT AAAAGAGAG GAAAAACCCC	840
	ACAAAGTATA ACAACCCCTT AAGATAATC TATTTTAA TGAATTAAT TTTTCAGTTT	900
35	ATACCATGG CCAATTACAA GATAAAATG TTAATTTT TAAAGATCC TTTGTTGACT	960
	TGCTTTTCA TCTCTTGCTA TTATATTTG TCACTTTAT TCAAGAAAT CTTATTTGCT	1020
	GAGGAAGGAC TTTCTGCACT TTAATGTTCT ACATCAAGA CTGGGAGGG TGGTGTTTAA	1080
40	CTTTTAAAAA AATGTTATTC TGATTAATAT AATATATG CTTTTTTTCA TGAAAGAGC	1140
	GCCACCTTGC AAGGTTTAGT GAGTTTATG GATTTTAA ACTTACGCG GAATTGCTGC	1200
45	TAGCTCCAAA AATTGCGAA GCAAAATCTA GCGGCAATG TTTTGAAGT TTGAAACTGA	1260
	TAAACAGATT TGCATTGAA GTGACTTCA ACATTAGTT CAGCATTAG TTAATAATAG	1320
	AAAGAGGAAT AAGACATCT YTTCTCTTA GAAAGATA CACTCAACT AATAATCCTT	1380
50	CCCACCTTCA TTGATCTCG CTGCTCTGA AACCTGATA GATGTGACA ATGATAAACA	1440
	TGATAATAGT GGTACTTTTG TAATTTTCT GGTGCAATTA AATAGTAGT AAAGCATGAG	1500
55	TTCAVCTTTT CTTCGAACAT YCCTATCTCT AGATGTAGT TACTTCAAT TGGGAATTAT	1560
	AAGTGTCTA ATTTTGTG TGTACCTGA TCGCCCTTT CTTTAAATAC CCACAGTGA	1620
60	ACAATTAAT ATCACTAT GACATATGA TTAAGTACA TATTTTAAAT ATAAATTTTA	1680



GGGGTAAATG TTTACTTCAA AATGACTCCA TATTTCAAAT ATCTGTITAG ACTGTGAAGG 1740  
 CCAAATAATT TTTAAGAAAA CATTGAAGA GTAGTGTGTT TGCATTGTG AATAATCTTA 1800  
 5 CTCACAGCRA GTAAACGTAA TAAAGCCRA CATTTAAGCC AAAAAAAAAA AAAAAA 1856

10 (2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1558 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

20 TGGGTATCCA TTCCTGNAAT TACTTTACTT AGGATAATGG CCTCCAGCTC COTCCAAGTT 60  
 GCTGCAAAAG GTATTATTTT GTTCCTTTTT GTGGCTGAGT AGTATTCCAT GGTGTATATA 120  
 TACCACATTT TCTTTATCCA CTCATTGCTT GATGGGCAGT TAGGTTGGTT CCACATCTTT 180  
 25 GCAATTGTGA GTTGTGCTGC TCCAGATATC ATCTTTAACT CCTTTGCCCT CTCCACATAC 240  
 ATTTCCAAGT CCTGTTCAAT CTACCTCCAA AATGTATCTT GTATCCATTC ATCTCTCTCC 300  
 30 ATCTTCAATC TATTTCAATG CCCCATCATC TCTTGCATGG AGGACTGTAA TAATTGGCTA 360  
 ACTGGCCTGT TCTTACATTT TAAATCAAA AGATGTGACA GGTGAAATGC CTATTTCAAT 420  
 GTCCATTGAT GGTTCGCTT ACACACCACC TGGCTGCCTG GTGTCCGAGT GGCAGAGTTG 480  
 35 AGCAGTGTGA AAAAGACTGC TTGGCCCTTT ACAGGGAAAG CAGGTCCACT GTGGCCTGTG 540  
 AGGACGAGAG CTCTGGGCAG GCTCGGACAC TGGCAGACCC TGGTCTGGC TGGCCAAGGC 600  
 40 AGCAGGGTAT GTGTTTCGGG TCACTCACAG GGCTCAGCAC CACTCCTCAT GGCTTCCTTA 660  
 CTGTTTCGGC AGAGGCTGAC CCGCGGCTGA TTGAGTCCCT CTCCCAGATG CTGTCCATGG 720  
 GCTTCTCTGA TGAAGCGGC TGGCTCACCA GGCTCCTGCA GACCAAGAAC TATGACATCG 780  
 45 GAGCGGCTCT GGACACCATC CAGTATTCAA AGCATCCCCC GCGTTGTGA CCACTTTTGC 840  
 CCACCTCTTC TGGTGCCCC TCTTCTGTCT CATAGTTGTG TTAAGCTTGC GTAGAAATTGC 900  
 50 AGGTCTCTGT ACGGGCCAGT TTCTCTGCCT TCTTCCAGGA TCAGGGGTTA GGCTGCAACA 960  
 AGCCATTTAG GGCAGCAAAA CAAGTGACAT GAAGGGAGCG TCCCTGTGTG TGTGTGTGCT 1020  
 GATGTTTCCT GGGTGGCCTG GCTCCTTGCA GCAGGGCTGG GCCTGGGAGA CCCAAGGCTC 1080  
 55 ACTGCAGCGC GCTCCTGACC CTTCCCTGCA GGGGCTACCT TAGCAGCCCA GCACATAGCT 1140  
 TGCCTAATGG CTTTCACTTT CTCTTTGTT TAAATGACT CATAGTCCC TGACATTTAG 1200  
 60 TTGATTATTT TCTGCTACAG ACCTGGTACA CTCTGATTTT AGATAAAGTA AGCCTAGGTG 1260

TTGTCAGCAG GCAGGCTGGG GAGGCCAGTG TTGTGGGCTT CCTGCTGGGA CTGACAAGGC 1320  
 TCACGAAGGG CATCCGCAAT GTTGGTTTCA CTGAGAGCTG CCTCCTGGTC TCTTCACCAC 1380  
 TGTAGTTCTC TCATTTCGAA ACCATCAGCT GCTTTTAAAA TAAGATCTCT TTGTAGCCAT 1440  
 CCTGTAAAT TTGTAAACAA TCTAATTAAA TGGCATCAGC ACTTTAACCA AAAAAAAAAA 1500  
 AAAAAAAAAA AAANRAAAAA AAAAGGGGGC CGCTCTAGAG GTCCAGTTA NGACGNGG 1558

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 948 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

TAAAAATCAT GCTCTGTACC ATCCTCACCG TAGTCATCAT CATCGCCGCG CAGACCACGA 60  
 GAACTACTGG GATCCCTAAA AACGCCCCCTG GTCCGGCCCC ACTCTGCGCC CCTCGATCTC 120  
 CCAGGCTCTT TCTGCAGWCA TACCGCGGAC CCAATGGGCG CCTGCGACAC CCGTTTCTGG 180  
 GCGCGTCAGA CTTGGATACA TCGTAAACTC CGCCTCCACG GAACGTCTCG CCKGCGAGC 240  
 AAGMTCGGAA TCCAGTTCTT CAGGAACCCC TCCAAAACCC ACACCCCCAG GGACGCCGCT 300  
 TTCCGGGATC CCGGSCAAAC GCGGACCCT CAGTCGCTCC AGCCCCCTC ACCCTCAAAG 360  
 TGTAGCGCCC CCAACCGAGC AACCTCGTT TGGTCCCTAA AACCCGCTT CCTCTATAAG 420  
 CACCGCCCCA GCTCTGACAA AACCCGCTT CCAGGTGGG AGGCTCCGCT TCTTTTCTTC 480  
 TCCGCGGGGT GATTCACTCC AGTGATTGGG TTTGTGGCTC CAGGCCTCCG CCACAGACGG 540  
 ACAGACCCCT CCCTTTCTTC CGGCAAAGG ACCGAGCCCT GGGTAGTAA GGGCCCCACA 600  
 CTCCTGTTTT TTGCAAGTAC ATTTTGTGCC YTCTCCACC CAGGTATCTG CCTATTTTCT 660  
 TGCTAATCCC AGAACCTTTC CTTTGTCTTT TTTAAGGAC ATTGGGAAG TTCCTGGTGT 720  
 AGGACCCCTC TCCCTGGGAT AAGAAACCTG CCTGTAAACG CTCTGTAAAT ACTCCCTTCC 780  
 ACCCATCCCA GCGCCTGGGC AGCGGGCAG AAGGGAATCC AGGCTATGGA CCTCCCAAGT 840  
 CCGCGCTCCC CGCTCCCTC GCGGGCCCCG CTTGTCTCTG ATCTGTGTGT GAGTGTGTGT 900  
 GAACTTCTGA AAGACAATAT TAAAGAGACT TAGTTGAAAA AAAAAAAA 948

(2) INFORMATION FOR SEQ ID NO: 55:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 990 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

10 GGGGAAGTGC AGTGACAGCA GGAGTAAGAG TGGGAGGCAG GACAGAGCTG GGACACAGGT 60  
 ATGGAGAGGG GGTTCAGCGA GCCTAGAGAG GGCAGACTAT CAGGGTCCCG GCGGTGAGAA 120  
 TCCAGGGAGA GGACCGGAAA CAGAAGAGGG GCAGAAGACC GGCGCACTTG TGGTTGTCAG 180  
 15 AGCCCCCTCAG CCATGTTGGG AGCCAAGCCA CACTGGCTAC CAGGTCCCTT ACACAGTCCC 240  
 GGGCTGCCCT TGGTTCTGGT GCTTCTGCC CTGGGGGCGG GGTGGGCGCA GGAGGGGTCA 300  
 20 GAGCCCGTCC TGCTGGAGGG GGAGTGCTG GTGGTCTGTG AGCCTGGCGG AGCTGCTGCA 360  
 GGGGGGCGCG GGGGAGCAGC CCTGGGAGAG GCACCCCTG GCGAGTGGC ATTTG/TGCG 420  
 GTCCGAAGCC ACCACCATGA GCTAGCAGG GAAACCGCA ATGGCACCAG TGGGGCQATC 480  
 25 TACTTCGACC AGGTCTCTGT GAACGAGGGC GGTGGCTTTG ACCGGGCGTC TGGCTCCTTC 540  
 GTAGCCCTG TCCGGGTGT CTACAGCTTC CGGTTCCATG TGGTGAAGGT GTACAACCGC 600  
 30 CAACTGTCC AGGTGAGCCT GATGCTGAAC ACGTGGCCTG TCATCTCAGC CTTTGCCAT 660  
 GATCCTGACG TGACCCGGGA GGCAGCCACC AGCTCTGTG TACTGCCCTT GGACCCCTGGG 720  
 GACCGAGTGT CTCTGCGCCT GCGTCGGGGG NAATCTACTG GGTGGTTGGA AATACTCAAG 780  
 35 TTTCTCTGGC TTCCTCATCT TCCCTCTCTG AAGGACCCAA GTCTTTCAAG CACAAGAATC 840  
 CAGCCCTGA CAACTTTCTT CTGCCCTCTC TTGCCCCANA AACAGCANAA GCAGGANANA 900  
 40 NACTCCCTCT GGCTCCTATC CCACCTCTTT GCATGGGAAC CTGTGCCAAA CACCCAAGTT 960  
 TAAGAAAAAA ATAAACTGT GGCATCTCCA 990

## (2) INFORMATION FOR SEQ ID NO: 56:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1603 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

50 GGTGACCCA CGCTCCCGC CCGCCGGCTC CGGAGCGGCT CTGCTTCCC GAGCGCGGA 60  
 CCGCGCCCTG GGGGAGGAGG GCGAACGACG CCGCGATGGC TCAGCGGCA CTCCCGGGT 120

	CCGCCGTCCT AGCCGCTGCT GTCTTCGTGG GAGGCGCCGT GAGTTCGCGG CTGGTGGCTC	180
	CGGACAATGG GAGCAGCCGC ACATTGCACT CCAGAACAGA GACGACCCCG TCGCCACGA	240
5	ACGATACTGG GAATGGACAC CCAGAATATA TTGCATACGC GCTTGTCCCT GTGTTCTTTA	300
	TCATGGGTCT CTTTGGCGTC CTCATTTCG CACTTCGCTT NAAGAAGAAA GGCTATCGTT	360
	GTACAACAGA AGCAGAGCAA GATATCGAAG AAGAAAAAGG TTGAAAAGWT AGRATTGAAT	420
10	GACAGTGTGA ATGAAAACAG TGACACTGTT GGGCAAATCG TCCACTACAT CATGAAAAAT	480
	GAAGCGAATG CTGATGTYTT AAAGGCGATG GTAGCAGATA ACAGCCTGTA TGATCCTGAA	540
15	AGCCCCGTGA CCCCCAGCAC ACCAGGGAGC CCGCCAGTGA GTCCTGGGCT TTGTCACCAG	600
	GGGGGACGCC AGGGAAGCAC GTCTGTGGCC ATCATCTGCA TACGGTGGGC GTGTGWTCC	660
	AGAGGGATGT GTGTCATCGG TGTAGGCACA AGCGGTGGCA CTTTATAAAG CCCACTAACA	720
20	AGTCCAGAGA GAGCAGACCA CGGCGCCAAG GCGAGGTCAC GGTCTTTTCT GTTGGCAGAT	780
	TTAGAGTNAC AAAAGTGGAG CACAAGTCAA ACCAGAAGGA ACGGAGAAGC CTGATGTCTG	840
25	TTAGTGGGGC TGAAACCGTC AATGGGGAGG TGCCGGCAAC ACCTGTGAAG AGAGAACCCA	900
	GTGGCACAGA GTAGCAGGTG AGCCGTGGTT TTGGTGACAT TGGGGGCAGA GTGGTGACGG	960
	GTGAGGAGAA GGTACTTGGA GCCTCCAGG TGCTGTGGCA GCATAGGAAT GGTATTTGAC	1020
30	AGGGAAGTGG GAGAGCTTTC CTTGACCCAG GAAGACTGAG GGGACTGAA CATGATTACT	1080
	TGTCTGCCTA GACCTTCTTG TAAAGAAGTC ACAAACTTAG TGCCTCCAGG GCCTTGGCTG	1140
35	TGTGATAATG AGGATAGAGG ATTACTTGTG AGGCAATGTG GCATGGTGGG GATTGTGGCA	1200
	AACTAGAATT CACATCACCC ACCATATAGG GCTTGCACTA CCACGAGGCA GAAAGCACCT	1260
	AGTGTGTCTG CATCTTCTTA CGCAAAAAAG ACAAATCCA GACTTCTAAA ATGTAAAATC	1320
40	ACTGATTTTC GATATTGGCA GCTTACTTTT TTTTTTTAAA CAACCATGCA GGCCAAATGA	1380
	CTTGTAATCT TGTCAACATT TTTAGGTAAA CTGTGACTTG AAAAAGTCTG GAGCAAACAA	1440
45	ACCAATGCTT TTTCCTTTTA TTCTGTGGG AACCAGTTT CTTTGTGTCA CAGTTTGTGAA	1500
	ACCTCAATAC GAATATTCTT CTTCCACCA AATATTTTGA GGCAATTGAA AACCCACAGT	1560
50	GATTTATTTT TGAATTGGC AATTTTAATT TTGCAAGACA ATT	1603

## (2) INFORMATION FOR SEQ ID NO: 57:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1052 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

5 TACAGCTCAG GATGCGCTGTA ACATTGTTCAT CTCTGGGCTT CTGGGTCTCTG CTTAGCCTGC 60  
 TTTTTCCTTG GAGGACTGAC CAGGGATGCG GCGGAGCAAC ATGTTACTAA ATCATACTCT 120  
 CCTCCCTACC TTTCCAGAC CTCTCACTCC TGCCTGGTGT TCCAACCGGT TCTGTGGCCA 130  
 10 GAGTATACAT TTTGGAACCT CTTGAGGCGC ATCTGCACT TCCAGATGAA CCATAGCCTG 240  
 CTTGAGCAGN AAGGCCCGAG ACATGTATGC AGAGGAGCGG AAGAGGCAGC ACCTGSAGAG 300  
 GGACCAGGCT ACAGTGACAG AGCAGCTGCT GCGAGAGGGG CTCCAAGCCA GTGGGGACGC 360  
 15 CCAGCTCCGA AGGACACGCT TGCACAACT CTGCGCCAGA CGGAAGAGC GAGTCCAAGG 420  
 CTTCTGCGAG GCGTTGGAAC TCAAGCGAGC TGACTGGCTG GCGGTCTGCG GCACTGCTATC 480  
 20 AGCCTGAATG AGGCTGGCCA CTTGCCACTT TGCCCTGCCC TCTGCCTCCA GGGCTCCMCT 540  
 MYCCTTCCTT TTCTTGGTGA AAGGCACCTC CTTCTCTGAT AATGAATGGT GTTCCCTTTG 600  
 CTTGGCTGGG GAGCCCCCCA GCGCAGGTTT GCTGGCCATA GATACCTTTG GGCTGCTGR 660  
 25 GACAGGCTCC TGAGGAGGAT TGAGGGTGAA AGTCTCCAC GAGTACACTA AACCTAGGTC 720  
 TGGTCACCAA TAGGGTTTGG AGAGCAAAGG GCCACAATC ATCAGCTGCC TGTCTCTTAG 780  
 30 ATGCACTTTT TTTTCCACC AGCACATCCT TCAACACACA GAATTTGAGG GAGAGTTCT 840  
 CCCCAAAACC CTAGCTCTTT ACCCTTCCAT TTTAGCCTTC CACCCAGCTT CCACAAAAGA 900  
 TTTGGCTCTA CCTTGGATCT GCTAGTAAAT AACTAATAGG CAGGCAGTTA TTTGGGTAAG 960  
 35 GAAAAAAGG GTGGGAGAGA CAGAAAATT GCCCACTGCT GCTCCTCCCC TTGGSTYTCC 1020  
 ACCTGGGATT TGCTATTGAA TCTTACCCT NN 1052  
 40

## (2) INFORMATION FOR SEQ ID NO: 58:

- 45 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 314 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

ACNCGNTGGC GGCGCTCTA GAACTAGGGG ANCCCCGGG CTGCAGGAAT TCGGCACGAG 60  
 55 CATAGACTTT TAAACTGGTA CGGTTCCTAG AGATGGTCTT TGGCCTCTG TTGTTGTGT 120  
 KGTTCCTTC TTTTCTTCT TCTCCTCTC CTTCTCTTC TCTCTCCTT CTTCTCTCT 130  
 60 TTTTTCCTCA GAGTCTTGCT CTGTCACCAA GACTGGAGTG AAGTGATGTG ATCTGGGCTT 240

ACTGCAACCT GGGAGGCAGA GGTTCAGTG AGTCGAGATG GTGCCATTGC TCTCGTTTGG 300  
 GCAACAGAG TGAACCTCTT GTCTCAAAA AAAAAAAAAA ATGAGGTTTA AGACAGTTTT 360  
 5 GTCATTACTG GTGGGATCTG GTCACACAAG ATAGCATTAA ACSTGACATG GCACATAAAA 420  
 TTGGTTAAAA AATTTTGTTC TTTAATTACG TAATGTAAAA GCCCAACAAA CACTTTATGC 480  
 AAGATTGGAA TGTATCTTCA AATTCAGATT TAATAACAT GTAAAGATCC TCTGTATATA 540  
 10 AAAGTTGTAT TTAATCCCTT GTGCCCCAAG AATGCTATAA AAGATCCCAA GAATGTTATC 600  
 TATGAAAAGA TAGCAATAGG GAATGGTGAA CAAATTAATT AATTGCCAA TTCTAAAAAA 660  
 15 CATGGACTTA AACCCCATGA AAAGTTGGTT CCATAGTTTT AACTGTTTTA TGGTTCCAAT 720  
 ACAAACCCAG AGTGGTTTAC ATTCCACAAT NACCAAATTT GCATCCAATN TTGGGGTAAT 780  
 TTINGGTATT TGCCATGGGA TACTATTCAT TTTT 814  
 20

## (2) INFORMATION FOR SEQ ID NO: 59:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1215 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AGAGGAAGTC TTTTGCCAAG CCTGTTCTCT GGAATAACGC CATCCAGGCT GGGAGGGGAA 60  
 35 GAGTGTCTCTG CTACACTCGT CCCCCTCCTG CCTCATCTTC CTTCTCAGCC TTGGTTCTCTG 120  
 ATGGGAACAG AATGGAGGGC CTGAGAACAT ACTTTCTAAA TGCCTTTGAC CCAGGAACCG 180  
 40 ATTATCTATA TTTGTTCCCA TTTTCCTTCA CCGTGACATT CCAGCATTGT CTGACTGTGA 240  
 GGTGGGCCTT TGAGAGCCTC CAGGTTCCCTC AAAACAGGCC TGAGCGATGG GCATCACACC 300  
 CTCTGCCTAC CCACRTGCCT GCTTACCTGC CAGATAACCA AGTGNAGATG TCTGCCAGTG 360  
 45 GCTAGTTTTC ACATTCTTAC TAGTGTGTTGG YTCACCTTTG GGCAAAGGCC CCCTCTAGGC 420  
 CTTGCCCCAC CTCCATCAA CGCAGACACT GTAGTCAGAC CTCAGYAATA TAGGAGGCAA 480  
 50 TAATCTTTTA ACAGTGTTTT GCAAACAAAC AAAACAGAGAA AAATCCCAGC CAGGGGAACT 540  
 CGCCACCTGC CCACGCTAGT TCCATCCAGC CTCAGACCC GCCCTTAGAC CAGGCAGGCA 600  
 AAGGCCCCCA TCACACTCGG CCACTAGTGG GTTCCTGAGG CCAAGAAAGA AACCAGACCC 660  
 55 TGTATGACAA GTTGGGKTCT TTCCAGAACA CGACAGAAC AGGGGGGGCC CCTTGTTAAT 720  
 GCCACTCCAT ACTCCAGAAG CATTATTCCT TATTTGGGAC AGCCAAGGCC AGATTACAG 780  
 60 GTTATTGTAG GAATPAAGAC TAGTTTACAA AGGARAAAGA GSCCCTGGAC TTCCCMAGGA 840

AAGGTCAGGT TAGGGCTCCT GTACCCATTC TGTTCACCA CTGTTTGATC TCTCTGGCCT 900  
 CCCACCAGGA ATGCCGTTTC CTTTTATGG ATCTGTTGGG AACCAGAGAG AATCAACAGA 960  
 TCAATGACAT AGGATCCGAA GTGCAATGAT AGTCACTTCT AGTTTGGCAT TTCACAACT 1020  
 CTGNACAGCA AGGTATTGGT AGGTACTCA ATTTCAAAAG GGCCCCATGG CCAAATATGT 1080  
 TTAGGAACCG CTGTTTGNAT TTCTTTTTTT GGAGACGCAT TGTATATAAT ATATGTCAA 1140  
 GGCTTTCGGA ATTCCTGCAG GAAAGAAATC AGCTTTGTTA AATCCNAAAA AAAAAAAAAA 1200  
 AAAAAAATAG ACTCG 1215

(2) INFORMATION FOR SEQ ID NO: 60:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 478 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: 'SEQ ID NO: 60:

ATTTCTTATG ACATGGGGGT TTGAATTGGT TGGCAATGT TTAATTTTAA TATCCATAAT 60  
 CAGTGAGGTC CTGCTGGCTG TAATCATTA TGTGAAATC TAAGGAGCTT AGTTCATGGC 120  
 TCTAGAATTT CACAGAAAAR TGYGMTATGA TACGAGCATT AAGTTTATTT CTTCTGATCT 180  
 TTGATGCAGC TTTGTTCACT TTATCTGTTT TTGTATTTAT TGGTCATCTA CTTCCCATGC 240  
 CAAAAGGGAC TGGTCTACAT AGCTGCGCTA AACACCTGAT CAAATCACTA AAAGAAAATG 300  
 TGTACCTCT ATGAATTAT CCTGATTGTA AGTTAAAAAT CAATATTTCC CCGTAGTGAG 360  
 GTTTGCTTTT TAAAAGAAK KCTTAAAAA AAAAAAAAAA AAACGAGTTN AAGAAAAGCA 420  
 AGCAAGCTCA GGTAAAGTGC ACACATTGGG CTAAGGAAGC TAGAGCCTGT GCAGANGC 478

(2) INFORMATION FOR SEQ ID NO: 61:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 618 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TATGACCTTG ATAACCCCAA GTTNGAAATT AACCTTCANT AAAGGGAACA AAAGCTGGAG 60  
 TTCGGCGCCT TGCAGTTGCA CACTAGTGGA TCCCAAGAA TTCGGGACGA GTCATAATGA 120

5 GCTACTAGGT AAGCCTTCTG GGACTTTCAG ATATTTTGGG GAAGATTGAT TTTTGTCTT 180  
 ACATGCTGTG GACCCCTGGC CATCAAATGG TATGGGAAG CTCATCCGTC TGTCTGTGAT 240  
 GGTCAATGCA GTCAGGCGTC TTTTGTAT TTTACTGGTG CTCAGTACTG TGCCAGATGC 300  
 TGTGGGAGC CGTGGTGGTA TGGAGGAGGA GTGCTCCAGA GGACTCTGCT GTGTGGCAGG 360  
 10 CCAGCATAAA CAAGCCAAGG GGAAAAGGCA GGCATGGAAT AAAGGGGGAG AATACCACTG 420  
 TGTGACTTAC TGCTGACTGT GTGGATTAGC CTATCAGCAG TAATCAAGCA GGGCGGAGGG 480  
 CATTATCTTT GAGCCAGAAG AGTGAGCACT GGSCCGAGGG TGGAGCATCA AGAGGGGGTG 540  
 15 TAGGACCAACA AGGCTTCTTN CNGGGGAGAC AACGTCAATA AGCNGTCAGT AGTCACCGAC 600  
 AGTTTTGGGA AGCAAGGG 618  
 20

## (2) INFORMATION FOR SEQ ID NO: 62:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 751 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

TCGACCCACG CGTCCGAGGA GCTGGACTTC TGAGACAGCC ATTCTCCTTG CATAGCACTG 60  
 35 TCTGCTGCTA CAGCTCATAG AAGTCAACAA TTTTCTTCAA CACTGGTAGG CAGCCTCTAA 120  
 ATGGCCCTGA TCACCCCTCAC CTCCTGCCAT TCACAGCQNT GTAAATTC ACCCCTGGAC 180  
 CTAGTGA CTC ACTTCTAACA ANGAGAATAC AGCAAAAGTA ACATCGCTTC TGAGGTGAGG 240  
 40 CTACAAGGAG ACTACGATGC CTGCCCTGGT CAGCCTTCTC CTGCTCTTTC CATTGCTCCC 300  
 TCTGATGGAA GCCAGTTGCC ATGTGATGAG GTGCCCTATG GAGAGGCCCA CGTGACAAGG 360  
 45 TATTGTAAAA AGCCTCTGAC CAATAGCCAT CTAGAAACGG AGGCCAGTC CAGCAGCCTC 420  
 TGAGATGAAT CCTGCCAACC TGAGCTTGGG GACAGATTCT CTCCTATTC TGCCCTGGGA 480  
 TGATCAGAGC CACCACCAAC ACCTTCACTG CCTGGTGAGA GGCCAAGCCA GTGAACCCAA 540  
 50 GGTAAACTGG ACAGATCCT GACCCACAGA AACTGAGATA ATGTTTGTTA TTTTAAGCTG 600  
 CTCAGTTTGT TACAGAGCAA TAGATACTA ACTCAAACAC CATAAATTC TAATATTTTA 660  
 55 TTCTATCACA CAAACCAGGT AATACCAAGT AATGCCATT ACTATACAGA TATTTTGTGA 720  
 ACACAATTAC ATGTGATTTT TTAAGAAGGC T 751  
 60



## (2) INFORMATION FOR SEQ ID NO: 63:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 780 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

CAGCAGTCA CAGTCCCGA TTCCCGGGTC GACCCACGGG TCCGGGTGG CAACTCCTGA 60  
 GGGCTGATG GTGACTTCA CATTTTCTA CCTCTCCTTC TATCTCTTC TAGAGCACCT 120  
 GGTATCCCA ACTTCTAGAC CTGCTCCAA CTAGTGACTA GGATAGAATT TGATCCCTA 180  
 ACTCATGTC TCGGTGCTC ATTGCTGCTA ACAGCATTGC CTGTGCTCTC CTCTCAGGG 240  
 CAGATGCTA ACCGGGCGAC GTCTTAATCC AACTGGGAGA AGCCTCAGTG GTGGAATTCC 300  
 AGGCACTTG ACTGTCAAGC TGGCAAGGGC CAGGATTGGG GGAATGGAGC TGGGGCTTAG 360  
 CTGGGAGTG GTCTGAGCA GACAGGGAA GGGAGAGGAG GATGGGAAT AGACAGTGGC 420  
 TGGTATGGT CTGAGGCTCC CTGGGGCTG CTCAAGCTCC TCCTGCTCCT TGCTGTTTC 480  
 TGATGATTC GGGGTTGGG ATCCCTTTG TCCTCATCTG AGACTGAAAT GTGGGGATCC 540  
 AGATGGGCT TCCTTCTCT TACCCCTTC CTCTCAGCCT GCAACTCTA TCCTGGAACC 600  
 TGTCTGCTT TCTTCCCA CTATGCACT GTGTCTGCT CCTCTGCAA GGCCAGCCAG 660  
 CTTCGGAGA GGAGAGAAAT AAACAGCAT TCTGATGCC AAAAAAAAA AAAAAAACC 720  
 GCGGCTGAA GTTATATCC CTTAAGTAA GGGGTTAAT TTAGCTTGG GCACTNGGCC 780

## (2) INFORMATION FOR SEQ ID NO: 64:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 588 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

TTCCGATTA ATCGACTCAC TATAGGAAT GCCGTGCCA TGACCCGGG TAACCAGCT 60  
 GAGCTGCCC GCGAGAGAA TATGAAAAG CAGAGCGACT CGGTTAAGG AAAGCGCCGA 120  
 GATGAGGGG TTTCTGCTG CCCCCGAG CAGAGCGACT CGGAGATCAT GCAGCAGAAG 180  
 CAGAAAAAG CAAACGAGA GAGGAGGAA CCGAAGTAGC TTTGTGGCTT CGTGTCCAAC 240  
 CCTCTGCCC TTGGCTGTG TCGCTGAGC CAGTCCACC ACGTCCGGT TTCTCTCTGT 300

322

AGTGCTCACA GGTCCCAGCA CCGATGGCAT TCCCTTTGCC CTGAGTCTGC AGCGGGTCCC 360  
 TTTTGTGCTT CCTTCCCCCTC AGGTAGCCTC TCTCCCCCTG GGCCACTCCC GGGGGTGAGG 420  
 5 CGGTTACCCC TTCCCAGTGT TTTTATTCC TGTGGGGCTC ACCCCAAAGT ATTAAGTA 480  
 CCTTTGTAAT TCCAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 540  
 AAAAAAAAAA AAAAAAAAAA AAAANNCGGG GGGGGGCCCC CCCCCCCC 588

## (2) INFORMATION FOR SEQ ID NO: 65:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 774 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

TTAAAGATG AAGAAATGAC AAGGGAGCGA GATGAGATGG AAAGGTGTTT GGAAGAGATA 60  
 25 ACGGGTCTRA GAAAGAAATT TAGGGCTCTG CATTCTPACC ATAGGCATTC TCGGGACCGT 120  
 CCTTATCCCA TTAAATTAAT TTCTCTGACA ATTCAATTAT TTCTGTTAT TAATGTTGCC 180  
 30 ACTGCTTTCT GTTGTCTGC ACTTTCTGA TAAATATTG CTATCGTTTT ACTCCAGTCA 240  
 TTCGATGTTG CTGAGATTTA CATATGACTC TTGTCAACAT CTCATCTTTT GACCCAATCT 300  
 TATTCATTTA ATAAGAGGTC TCATTCAATT GCATGGAAAA ATGCTCATTG TATATTGCAA 360  
 35 AGTGAAAATA ACGAGTTGCA AAACAGTGTA TACATATATG TGTGTATATA TGTACACTTT 420  
 ATTTGTACAT TTCTATGTA CATAATGCAA AGGAAAGTGT CTGATTTTAT TATACACCRA 480  
 40 AGGTTAACAG TGAATCTCTG TGTGATCTCT TTTTCTTCT TTTTGCCTAT CTGCATCTTC 540  
 TCACTTGCCA AAAAATGAAT ATATGTTTAT GTGTGTATAT TACTTGTGTC ACAA AAAACC 600  
 CTAAAGTAGA CAGTAAAAGA ACTTGTCAAT CGCCTTTGGA AGGCAATGAA ACACTTAATA 660  
 45 AACTCTCAAT AACAGAAGCG TAAAAATGAA ATGTAAACCT CCAATTACCT CTGGATCTCT 720  
 TAGCCAGAGT AATAAACTGG TAATTATTAC AGATAAAAAA AAAAAAAAAA AANA 774

## (2) INFORMATION FOR SEQ ID NO: 66:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1866 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

	ACCCACGCGT CCGGTGCTCT TCTTCAGCAC ATGCCAAAGC TGTTCCCTCAC GGCCTGTGAG	60
5	ACAAGAGCAT CTTGGATGTA GCACAATGGA AGAGTTAGAT GCCTTATTGG AGGAACTGGA	120
	ACGCTCCACC CTTCAGGACA GTGATGAATA TTCCAAGCCA GCTCCTCTTC CCCTGGATCA	180
10	GCATTCCAGA AAGGAGACTA ACCTTGATGA GACTTCGGAG ATCCTTTCTA TTCAGGATAA	240
	CACAAGTCCC TTGCCGGCGC ATTCGTGTAT ACTACCAATA TCCAGGAGCT CAATGTCTAC	300
	AGTGAAGCCC AAGAGCCAAA GGAATCACCA CCACCTTCTA AAACGTCAGC AGCTGCTCAG	360
15	TTGGATGAGC TCATGGCTCA CCTGACTGAG ATGCAGGCCA AGGTTGCAGT GAGAGCAGAT	420
	GCTGGCAAGA AGCACTTACC AGACAAGCAG GATCACAAGG CCTCCCTGGA CTCAATGCTT	480
20	GGGGGTCTSG AGCAGGAATT GCAGGACCTT GCCATTGCCA CAGTGCCCAA GGGCCATGT	540
	GCATCCTGCC AGAAACCGAT TGCTGGGAAG GTGATCCATG CTCTAGGGCA ATCATGGCAT	600
	CCTGAGCATT TTGTCTGTAC TCATTGCAAA GAAGAGATTG GCTCCAGTCC CTTCCTTGAG	660
25	CGGAGTGGCT TGGNCTACTG CCCCACCGAC TACCACCAAC TTTTCTCTCC ACGCTGTGCT	720
	TACTGCGCTG CTCCCATCCT GGATAAAGTG CTGACAGCAA TGAACCAGAC CTGGCACCCA	780
30	GAGCACTTCT TCTGCTCTCA CTGCGGAGAG GTGTTTGGTG CAGAAGGCTT TCATGAGAAG	840
	GACAAGAAGC CATATTGCCG AAAGGATTTT TTAGCCATGT TCTCACCCAA GTGTGGTGGC	900
	TGCAATCGCC CAGTGTGTGA AACTACCTT TCAGCCATGG AACTGTCTG GCACCCAGAG	960
35	TGCTTTGTTT GTGGGGACTG CTTCAACAGT TTTTCTACTG GCTCCTTCTT TGAAGTGGAT	1020
	GGACGTCCAT TCTGTGAGCT CCATTACCAT CACCGCCGGG GAACGCTCTG CCATGGGTGT	1080
40	GGGCAGCCCA TCACTGGCCG TTGTATCAGT GCCATGGGGT ACAAGTTCCA TCCTGAGCAC	1140
	TTTGTGTGTG CTTTCTGCTT GACACAGTTG TCGAAGGGCA TTTTCAGGGA GCAGAATGAC	1200
	AAGACCTATT GTCAACCTTG CTTCAATAAG CTCTTCCCAC TGTAATGCCA ACTGATCCAT	1260
45	AGCCTCTTCA GATTCCTTAT AAAATTATAA CCAAGAGAGC AGAGGAAAGG GTAAATTTTC	1320
	TGTTACTGAC CTTCTGCTTA ATAGTCTTAT AGAAAAAGGA AAGGTGATGA GCAAATAAAG	1380
50	GAACTTCTAG ACTTTACATG ACTAGGCTGA TAATCTTATT TTTTAGGCTT CTATACAGTT	1440
	AATTCTATAA ATTCTCTTTC TCCCTCTCTT CTCCAATCAA GCACCTGGAG TTAGATCTAG	1500
	GTCTTCTAT CTCGTCCCTC TACAGATGTA TTTTCCACTT GCATAATTCA TGCCAACACT	1560
55	GGTTTCTCTA GGTTTCTCCA TTTTCACCTC TAGTGATGGC CCTACTCATA TCTTCTCTAA	1620
	TTTGGTCTG AACTTGTATT CTTTCACGT TTTCCCATTT CCGTGTGGCT CACTGTCTTA	1680
60	CAATCACTGC TGTGAATCA TGATACCACT TTTAGCTCTT TGCATCTTCC TTCAGTGTAT	1740

TTTTGTTTTT CAAGAGGAAG TAGATTTTAA CTGGACAACCT TTGAGTACTG ACATCATTGA 1300  
 TAAATAAACT GGCTTGTGGT TTCAATAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1360  
 5 AAAAAA 1365

10 (2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1152 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

20 CTCAAGGATG TAAAGGCTCT GCAGATTTCG GGAGGCCTGT CTCCCAGCAC CTGATGGGAC 60  
 ACTTTTGGCC CCACTGTAAA TTCTGGGTGT ATCCTCCACT GTATGCTGTC ACCCCAAGGG 120  
 CAAGCACTGC ATCTGCTTAG TGAAGGATTT ATTGTTCCGA AGATACATTT TCCCTTTKAG 180  
 25 CAGAGAGTGG CGTATCCTGG CAGTCTTCGG TGAGCCAGTT GTACCAGGAT TATGAAATGC 240  
 AGATGTTTAC TGTGTCATTG TTGCTGTCAT TGCTACTGAG GAGTACTGAC CAGAATCATC 300  
 30 TGCAACTYTT AGTTGGCAGA GAGGACCACT ATGCCGGGTA GCTCTTTTCT TTCTTGCCAT 360  
 TGTGGGGATG ATTCCAGGCC AAAGATGATG GAAAGTATG GAAATCATCT GAAAGGTTGA 420  
 AGCTTGGCAC GTGAAGCCAT TCATGACTTT GTAAGGCAGT TTTGCTGAAG GCCAGTTCTG 480  
 35 CCCTGGGAGG GACGGAGGTG AATCCTCCTG AGTACCTGTG GTTTTCTTAC TTCCTGCTGA 540  
 ATTTACCTAA GTGCCTGTG TTGCTTGCT GTGGAGGCTT TCTGGTATTT CATTTTCAGGT 600  
 40 GCAGATGCCT TCACTTTGCC ACCRAAAAAA CCCCMACCAA ACCTAAGACC TTAAGTCAAC 660  
 TAAGTYTNCC AAGTACTTTT TAACCCAATG GGATGAACAG CCTGTGGTCT GCTCAGATCA 720  
 CCCTGAGTGC GTGTGAGAAG GCMTNGGCTT TGCCAGGAAA TCCAGGAAGG CAGGGCCGGG 780  
 45 CTGTGTTGGA AGCTGGCTTA GCTGGTGGGG CAGCCTTATT TCAATTAAAA GGGCATTGAC 840  
 TGGGAGCAGC AGTCTGGAG TTTGTTGCAT TTCCTATTGC CCTCAAAATG AGAAACCAGG 900  
 50 AAAATAGCAG ATTGGAGCCT TCGAGAAGGC AGTAAATGCC TGTTTTATT GACAAAAGGA 960  
 AAACATTTTA CTGCCATCTC ACTGATGGCA TCTCACTGAC TTAAATGAA GGCANGTTGT 1020  
 AGTAAAAAAA AAAGTCTACA TTTTCCACC GCCACGTTCT TATATCCTGT TTGTCAGCCA 1080  
 55 CTGCTCANAA GGGCATGTTG TCTTGGGGAN TANAGGCCCT CTCCTTCCCT CGTTTTCCCT 1140  
 ATAGGTTGGG TG 1152

## (2) INFORMATION FOR SEQ ID NO: 68:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2483 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

```

5      ACCAGGCGGT GCGCTGGGGG CCGGAGCAGC GCGKAGCCCG GCTCGGCCAC ACCGATCGCC      60
15     CGCCGCCATG GGCTCCTCGC AAAGCGTCGA GATCCCGGGC GGGGGCACCG AGGGCTACCA      120
      CGTTCTGCGG GTACAAGAAA ATTCCCCAGG ACACAGAGCT GGTTTGGAGC CTTTCTTTGA      180
20     TTTTATTGTT TCTATTAATG GTTCAGATT AAATAAGAC AATGACACTC TTAAGGATCT      240
      GCTGAAASCA AACGTTGAAA AGCCTGTAAA GATGCTTATC TATAGCAGCA AAACATTGGA      300
      ACTGCGAGAG ACCTCAGTCA CACCAAGTAA CCTGTGGGGC GGCCAGGGCT TATTGGGAGT      360
25     GAGCATTGCT TTCTGCAGCT TTGATGGGGC AAATGAAAAT GTTTGGCATT TCGTGGAGGT      420
      GGAATCAAAT TCTCCTGCAG CACTGGCAGG TCTTAGACCA CACAGTGATT ATATAATTGG      480
30     AGCAGATACA GTCATGAATG AGTCTGAAGA TCTATTCAGC CTTATCGAAA CACATGAAGC      540
      AAAACCATTG AAATGTATG TGTACAACAC AGACACTGAT AACTGTGAG AGTGATTAT      600
      TACACCAAAT TCTGCATGGG GTGGAGAAGG CAGCCTAGGA TGTGGCATTG GATATGGTTA      660
35     TTTGCATCGA ATACCTACAC GCCCATTTGA GGAAGGAAAG AAAATTCTC TTCCAGGACA      720
      AATGGCTGGT ACACCTATTA CAGTCTTAA AGATGGGTTT ACAGAGGTCC AGCTGTCTC      780
40     AGTTAATCCC CCGTCTTTGT CACCACCAGG AACTACAGGA ATTGAACAGA GTCTGACTGG      840
      ACTTCTATT AGCTCAACTC CACCAGCTGT CAGTAGTGT CTCAGTACAG GTGTACCAAC      900
      AGTACCGTTA TTGCCACCAC AAGTAAACCA GTCCCTCACT TCTGTGCCAC CAATGAATCC      960
45     AGCTACTACA TTACCAGGTC TGATGCCTTT ACCAGCAGGA CTGCCCAACC TCCCAACCT      1020
      CAACCTCAAC CTCCCAGCAC CACACATCAT GCCAGGGGTT GGCTTACCAG AACTTGTAAG      1080
      CCCAGGTCTG CCACCTCTTC CTTCCATGCC TCCCCGAAAC TTACCTGGCA TTGCACCTCT      1140
50     CCCCCTGCCA TCCGAGTTCC TCCCGTCATT CCCCTTGGTT CCAGAGAGCT CTTCTGCAGC      1200
      AAGCTCAGGA GAGCTGCTGT CTTCCCTCCC GCCCACCAGC AACGCACCT CTGACCCTGC      1260
55     CACAACACT GCAAGGCAG ACGCTGCCTC CTCACCTACT GTGGATGTGA CGCCCCCAC      1320
      TGCCAAGGCC CCCACCACG TTGAGGACAG AGTCGGGAC TCCACCOCAG TCAGCGAGAA      1380
60     GCCTGTTTCT GCGGCTGTGG ATGCCAATGC TTCTGAGTCA CCTTAACCTT GAACCATCT      1440

```

326

TTGGAATTGG CGTGGTATAT TTAACCACGG GAGCGTGTCT GGAAACGCAA ACTATCATT 1500  
 ATTTTCATACT AGTTTGTACC GTATCTGTAG GCATCCTGTA AATAATTCCA AGGGGAAAAC 1560  
 5 TAAACGAGGA CGTGGGTGTG ATCCTGCCAG GTTGAGTGGG GCTCACACGC TAGGGTGAGA 1620  
 TGTGAGAAAG CGCTTGATTT TTAACAACCC AAAAAGAATT GTAAGGGTGG CTTGCTGCCA 1680  
 GGCTTGCACCT GCCGTTCCTG GGGGTGTGCA TCTTCGGGAA AGGTGGTGGC GGGGCGTCCA 1740  
 10 CTAGGTTTCC TGTCCCCTGC TGCTCCTTCC GTAAGAAAAT GAAATATTCT ATGCCTAATA 1800  
 CTCACACGCA ACATTTCCTG TACTTTGTAA GTCGTTTGGC AGAATGCAGA CCACCTCACT 1860  
 15 AAACGTGAAA CGGTAAAGAG ATTTTACTTT TTGGTCTCCG TGAGTCCCAT CTCTACTAAG 1920  
 GTTTACACAG GAATTCACCC TGAAGACTTG TGTAAAGTT CTACAGCGCG CACTGTTAAC 1980  
 TGAACGTCTT TTTCTTCAGC CTATACCGCG ATCCTTGTTT TGAGCTCTCA GAATCACTCA 2040  
 20 GACAACATTT TGTAACGTCT GCTGTGTCTT TCTACATACA CCTTATAAAG TGACATTTCA 2100  
 AAAGAAATAA GGTGCCACAG TTTTAAACCA GAAGGTGCCA CTCTGTGGCT CCTTGTAGTA 2160  
 25 TTATAGCTAT ACTGGGAAAG CATAGATACA GCAATAAAGT ACAGTAATTT TACTTTTTTT 2220  
 CTTGTGTAC ATCTAAATTA CAACCTTAA TTGCCACGTG TGCCTTACT ACTCTCCAGT 2280  
 ATGTCTTATT ACTCTCCAGT ATGTCACGCA TCTTTAACTT TTCACGTCTT ATGTTTGCTT 2340  
 30 TCTCCCATTT TTAAGAGATG GTAAGTTAAC TGAATTGAT TTAAGTAATG AAATTAATG 2400  
 CAGATATCCC TGTTTTGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 2460  
 35 AAAAAAAAAA AAAAAAAAAA AAA 2483

40 (2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 536 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

50 GAGAAATGGA GCTTTGTTAG ATAAAAATTT TTTCAACGCA AACAGTCATT TTCCAGTGAA 60  
 AGGAGAGCGT ATCCGCCGTA GATGGACTT AGATCGTGTA AAAGCTGAGG CCACCGAGGA 120  
 TATAACCTCC GGGGTCCCTT GCCTCCTTTT CCTTAGACTC CCTCCAACT CGTGTATCTT 180  
 55 TCCTTCAGCA GTACTGGGCT CCACGCGAAC CTAGTCCTTT GTCTTTACCC TATTACCTTT 240  
 CATAACATCC TAGTTGAAAA GTATTATTC AACCGGTTT GAAATGAGA ACXGGTTCAC 300  
 60 AGARGCTAGG TTAAGTGGCA AGTTCGTTCA ATTAGTAACC ACTAACGCCA GCACTGCCAG 360

TTTCTTGCTT CCGAATTCTC ATGGTAGCTT TCACCARGCT CCCCCTQMAA TGCTAACGTC 420  
AACTACTGAA CTAGATTAGC AAAAAGGTCT TTTAACAGAA TTCTTGTTT TCAGAGAGAG 480  
5 TTTCTTTTAT GAAGCGCCCC ATTCTACAG AGGAAAATAA ACTCCAAGCA GCCAGT 536

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 865 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

20 CCACGCTCC GGCCTTTCTT GGCCAGAGGC GCCGGTTGGA CTCACGGGCG GGGCATGATG 60  
GGTAACAGGA CCGGTGGGGT CCCCAGGAAG TCCTAGAGGG GGTGGGGTT TGGGTGACA 120  
25 AGCTTTCCTC GTCTCTCCC GACAGAGCTG ACGTGTCTG GGTTCACCG GGAGCGGCA 180  
TTTCACCGG ACGGAGGGT TCGGGGTGTC CCGGGCTGGG CAATACGTAG GGGTTGCCG 240  
GCGGTGTGGG GAGTTGGGGC GTGTGGCTGC AGTCCCGGA GTTCTGGAG GGGTCCGCC 300  
30 CACCGAGCTT CCGGACCGG TGATCTGCC GTAGCTGCC GGANGGARG CGGAGCTGAC 360  
TCTCCGTCCC TTCTCCATC CCTCCAGTG GTGGGTACGG GCACCTCGCT GCGCTCTCC 420  
35 TCCCTCTGT CCTGTCTGT CTTGTCTGG ATGCAGATGT ACAGCCGTCA GCTGGCTCC 480  
ACCGAGTGGC TCACCATCA GGGCGGCTG CTTGGTTGG GTCTCTCGT GTTCTCGCTC 540  
ACTGCCTTCA ATAATCTGA GAATCTGTC TTTGGCAAAG GATTCCAAGC AAAGATCTTC 600  
40 CCTGAGATTC TCCTGTGCT CTTGTGGCT CTCTTTGCAT CTGGCTCAT CCACCGAGTC 660  
TGTGTACCA CTGTCTCAT CTTCTCCATG GTTGGTCTGT ACTACATCA CAAGATCTCC 720  
45 TCCACCCTGT ACCAGGCAGC AGCTCCAGTC CTCACACCAG CCAAGGTCAC AGGCAAGAGC 780  
AAGAAGAGAA ACTGACCTG AATGTTCAAT AAAGTTGATT CTTGTAAAA AAAAAAAAAA 840  
50 AAAAAAAAAA AAAAAAAAAA AAAAA 865

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 932 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

5 TCATCATATA CAAAGTTTTT CGTCACACTG CAGGGTTGAA ACCAGAAGTT AGTTGCTTTC 60  
 AGAACATAAG GTCTGTGCA AGAGGAGCCC TCGCTCTTCT GTTCTTCTC GGCACCACCT 120  
 GGATCTTTGG GGTTCCTCAT GTTGTGCACG CATCAGTGGT TACAGCTTAC CTCTTCACAG 180  
 10 TCAGCAATGC TTTCCAGGGG ATGTTCAATT TTTTATTCCT GTGTGTTTTA TCTAGAAAGA 240  
 TTCAAGAAGA ATATTACAGA TTGTTCAAAA ATGTCCCCTG TTGTTTGGGA TGTTTAAGGT 300  
 15 AAACATAGAG AATGGTGGAT AATTACAACG GCACAAAAAT AAAAATTCCA AGCTGTGGAT 360  
 GACCAATGTA TAAAAATGAC TCATCAAATT ATCCAATTAT TAACTACTAG ACAAAAAGTA 420  
 TTTTAAATCA GTTTTTCTGT TTATGCTATA GGAAGTGTAG ATAATAGGT AAAATTATGT 480  
 20 ATCATATAGA TAACTATGT TTTTCTATGT GAAATAGTTC TGTCAAAAAT AGTATTGCAG 540  
 ATATTGGAA AGTAATGGT TTCTCAGGAG TGATATCACT GCACCCPAGG AAAGATTTTC 600  
 25 TTTCTAACAC GAGAAGTATA TGAATGTCTT GAAGGAAACC ACTGGCTTGA TATTCTGTG 660  
 ACTCGTGTG CTTTGTAAAC TAGTCCCCTA CQACCTCGGT AATGAGCTCC ATTACAGAAA 720  
 GTGGAACATA AGAGAATGAA GGGGCAGAAT ATCAAACAGT GAAAAGGGAA TGATAAGATG 780  
 30 TATTTTGAAT GAACTGTTTT TTCTGTAGAC TAGCTGAGAA ATTGTTGACA TAAAATAAG 840  
 AATTGAAGAA ACACATTTTA CCATTTAAAA AAAAAAATAA ACTNGAGGGG GGCCCGGTAC 900  
 35 CCAAATCCCC GCATAGTGAT CGTAAACAAT CT 932

## (2) INFORMATION FOR SEQ ID NO: 72:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 996 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

50 CGCCTGGCAC CATGAGGACG CTTGGGCTTC TGCCTGTGCT GCTGCTGCTC CTGGCCGGAG 60  
 CCCCCGCGGC GCGGCCCACT CCCCCGACCT GCTACTCCCG CATGCCGGCC CTGAGCCAGG 120  
 AGATCACCCG CGACTTCAAC CTCCTGCAGG TCTCGGAGCC CTCGGAGCCA TGTGTGAGAT 180  
 55 ACCTGCCCCAG GCTGTACCTG GACATACACA ATTACTGTGT GCTGGACAAG CTGCGGGACT 240  
 TTGTGGCTTC GCCCCCGTGT TGGAAAGTGG CCCAGGTAGA TTCCTTGAAG GACAAAGCAC 300  
 60 GGAAGCTGTA CACCATCATG AACTCGTTCT GCAGGAGAGA TTGGGTATTC CTGTTGGATG 360



ACTGCAATGC CTTGGAATAC CCAATCCCAG TGACTACGGT CCTGCCAGAT CGTCAGCGCT 420  
 AAGGGAAC TG AGACCAGAGA AAGAACCCAA GAGAACTAAA GTTATGTCAG CTACCCAGAC 480  
 5 TTAATGGGCC AGAGCCATGA CCGTCACAGG TCTTGTTGTTA GTTGTATCTG AACTGTTAT 540  
 GTATCTCTCT ACCTTCTGGA AAACAGGGCT GGTATTCCTA CCCNGGAACC TCCTTTGAGC 600  
 10 ATAGAGTTAG CAACCATGCT TCTCATTCCTC TTGACTCATG TCTTGCCAGG ATGGTTAGAT 660  
 ACACAGCATG TTGATTGGT CACCTAAAAA GAAGAAAAGG ACTAACAAGC TTCACTTTTA 720  
 TGAACAACTA TTTTGACAAC ATGCACAATA GTATGTTTTT ATTACTGGTT TAATGGAGTA 780  
 15 ATGGTACTTT TATTCTTTCT TGATAGAAAC CTGCTTACAT TTAACCAAGC TTCTATTATG 840  
 CCTTTTCTA ACACAGACTT TCTTCACTGT CTTTCATTTA AAAAGAAATT AATGCTCTTA 900  
 AGATATATAT TTTAYGTAGT GGTGACAGGA CCCACTCTTT CATTGAAAGG TGATGAAAAT 960  
 20 CAAATAAAGA ATCTCTTCAC ATGAAAAAA AAAAAA 996

25

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 785 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

40

45

50

55

60

GGCACGAGGG GCTTTGCGTA CACAATAGCT GCTAGGAGTA CCCAAAGCCT GARTACARCC 60  
 TGCTGGTCTC ATGGCCACGT GTGAGCAGGC CAGCGTCAMA CGGCTCGGTG TGACCCGTCC 120  
 CGRAGACTGA AATGGGCCTG GGTCTTCTCC TKGTCCTGTG ATWAAAGTCC TCTCTGAAA 180  
 GTGGAGAGCA AAGGCACACA GAGGTGCGCG CTCACAAGAA TTCCTCCCGG TGACTGGGTA 240  
 ATCAATGTTA CTGCTGTTTC CTTTGCAGGA AAGACCACAG CAAGATTCTT TCATTCTGCT 300  
 CCTCTAGCC TGGGGGACCA GCTCGAACT GACCTCGGAC ATCAAAGGAG GGATTATGTG 360  
 GGTGCTAAAG CCATCGGCCC ACAGCCCTGT TCACRTCTTG GTGCTTCTCT TTCCAGAGG 420  
 CTGGTCCCAG CCAGGCACAC ACAAAGGCA GATTCTCGTA AACSCAGCCT CCCTCCCTGG 480  
 AGGCTGCCCTG CTGCCCTGGA TCTGGAGTGG AGCTGCTCTG AGATTTTGAG TTCTTCTGCA 540  
 GAGATGATTA AATATATCCA AGAGACATTG GAAAACCTGC TGAACATTTT ACATTGGTCT 600  
 GCTCAGCACA TGGCTGGATG CGGATATTTC TATAATTCCA GAAAGTCACA CAGCTCCTCT 660  
 GTATGAGACC AGTGGGCGCC ATTTAAAAGA ACAGGATGAG AATCTAAGAT ATATTATTAA 720  
 60 TAAATGTAAT GGATTTTTTT TTTGTAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 780

AAAAA

785

5

(2) INFORMATION FOR SEQ ID NO: 74:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1069 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

20

TCCTCACCAT TCCCTAGGN CAGGTCCTG CAGGTCCCAC ACTTCTCCCA GGTCCCTAAA	60
CTTGGGTCGG TCCTTTCCCT GGAGTAGCTG GNTCCTCCAG TCGAGGTCCC TGTTCAGTCG	120
GTTCTTAGGC TCCTGCACAT GAAGGTGTGT GCCTGTGGTG TGTGGGCTGC TCTAGGAGCA	180
GATACAGGCT GGTATAGAGG ATGCAGAAAG GTAGGGCAGT ATGTTTAACT CCAGACTTGG	240
CACATGGCTA GGGATACTGC TCACTAGCTG TGGAGGTCTT CAGGAGTGGA GAGAATGAGT	300
AGGAGGGCAG AAGCTTCCAT TTTTGTCTTT CTAAGACCC TGTATTGTGT GTTATTTCCT	360
GCCTTTCCGA GTCTTGCACT GGGCTGCCCT GTACCTGAA CCTCATGAGC CTCTAAGGGA	420
AAGGAGGAAC AATTAGGACG TGGCAATGAG ACCTGGCAGG GCAGATACA AGCCCAGCAC	480
CAGTGTCCCA GCCTTACTGG GTCTTACCC TGGGCCAAAC AGGGAGGGCT GATACCTCCT	540
TGCTCTTCCT AGATGCCAC CTCTACAAT CTCAGCCCAC AAGTCTCTC CACCCTAGGG	600
GGCTTGCTGC ATGGCAATAA CTCATAATCT GATTGGAGG TTTGCCCTTT ACAGGGGCAG	660
ATTTTCTGCT CAGTTCAACA ATGAAATGAA GAGGAACTCC CTCTTCTAC AGCTCACTTC	720
TATCAGAGGC CCAGGTGCCT CAGAGCCACA TTGAGTTGCT TTTTCTGGGA TGAGGAAGTA	780
GGTTAAACT CCCCAGTTTC CTGAGGGAGG CTCCTGACAG GTGCCCTTTG TCAGACCCTA	840
CCACAGCCTG GATAGGCAGC CACATTGGTC CTCGCCCTTG CTCGNACTC CGTGGTGGTC	900
CTGCCCTTCT CCCTGCATGC CTGTGGGTCT GCTCTGGTGT GTGAAGGTCT GTGGGTTAAC	960
TGTGTGCTTA CTGAACCTGG CAAATAAACA TCACCCTGCA AAGCCAAAAA AAAAAAAAAA	1020
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1069

55

(2) INFORMATION FOR SEQ ID NO: 75:

60

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 831 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

5  
GGCATTAGA TCACTGTGGA CCTAAAACAA ACRAACAACT ATAAGGAAAA TGCCATTAGA 60  
AATGGTCTGG GGATCAGTTT ATCACTGCAG TTGTTACATC ACCCCATGGT CTAAAATACA 120  
10 GAGCTTTAGT CTGTCTCTGT TTCAGTTCAT TTTACAGGAG GTGAACATCA CACTTCCAGA 180  
AAACTCTGTC TGGTATGAAA GGTATAAATT TGATATTCCT GTCTTTCACT TGAATGGCCA 240  
GTTTCTGATG ATGCATCCAG TAAACACCTC AAAACTTGAA AAACAGCTCC TGAAACTTGA 300  
15 GCAGCRAAGT ACTGGARGCT GACTGATGCC CTCATGATTT TCCACCCTCT CTTCCTATAA 360  
AGCATCTTCC TAAGGAAATG AMCATGGCCT GATACTCATT TTGTCATTG TACAGAGCCC 420  
20 TAAGGATGTT CTGAATTCAG TGGTGCCAAA TAAATGTTGA CATTCCCCTT TTGGTTGATG 480  
GAAGTATCAG TGTGGGAACT GTTTGCTTAA TGGCATTTTA TAAAATAAKA AKAKCATATT 540  
AGCAGGGAGG GAGATGATGG AGGGAGGGAG AAGTCCATTT GTCTTATTTA TCCTTTTTGT 600  
25 ATTAATAGAG AAGCACTTCA CAGTCACTGG CAATGCCATT TATAGGAAGA AGGTTCTGCA 660  
TTCCTGCTGC TCCCGGAGGG CTAACTTTT TAATGAAAGA ATAAATGCTC TTCCACTCAG 720  
30 TAGATAAAGT GAAATGTGAA TTGTTAATAA CTGTGCACGG TCAATAAAGC GATGTTTTAA 780  
CGAATACAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAACTCG A 831

35

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:

40 (A) LENGTH: 590 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

TATATATAGA CNGTTAATAG TCGTGANTGN TGTGNACGAA CATTAACGGA AGTAGCATGT 60  
AGCCAGTCCA ATAACNTATA AGGACAAAGT GGAGTCCACC CGTGCGGCGG TCTAGACTAG 120  
50 TGATCCCCC GGCTGCAGGA TTCGGCACGA GCTGCCAGGT GAGGAGCAGA GAGACTGTTC 180  
CCTTGGGTGG AGAGGTGTGG GCATGAGAGC CACCCATTGC CAAGCAGCPA GAATGTTCTG 240  
55 GCTTTTTTCC CTTCCTAAAT ATGCAGGGCT CAGGCTCCCA ATTCCGGGGC TGTCTGCTTT 300  
GCTTGTGTTT CTCTGTCCG TGTTCTCCCG GAGGGCCCG GTCGAATCA CGACAGGGAG 360  
CGAGACGCTT CCCAAAAAC TGCAGGGCTA TTTCCAGAA TTTGGTTTTT AAGTACAAA 420  
60

CTTTTGTCC TGTAAGATAT ATGCAGCCTC ACAGAAGCAG CCTCTGCCCTG CACTTTACCA 480  
 GCTACGTTTT TATCTTAAGC ACATGGGGCT CCCTTAGAAC TTA CTCCACT GATTTAAAAA 540  
 5 AAAAAAAAAA AACTCGAGG GGGGGCCCCG TACCCATTCT CCCTAAAAGT 590

10 (2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1274 base pairs

(B) TYPE: nucleic acid

15 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

20 GAGCCACCAC ACCTGGCCTG GAAGGAACCT CTAAAATCA GTTTACGTCT TGTATTTTGT 60  
 TCTGTGATGG AGGACACTGG AGAGAGTTGC TATTCAGTC AATCATGTCT ACTCACTGGA 120  
 CTCTGAAAAT CCTATTGGTT CCTTATTTT ATTTGAGTTT AGAGTTCCCT TCTGGGTTTG 180  
 25 TATTATGTCT GGCAAATGAC CTGGGTATC ACTTTTCCTC CAGGGTTAGA TCATAGATCT 240  
 TGGAAACTCC TTAGAGAGCA TTTGCTCTT ACCAAGGATC AGATACTGGA GCCCCACATA 300  
 30 ATAGATTTCA TTCACTCTA GCCTACATAG AGCTTTCTGT TGCTGTCTCT TGCCATGCAC 360  
 TTGTGCGGTG ATTACACACT TGACAGTACC AGGAGACAAA TGA CTTACAG ATCCCCCGAC 420  
 ATGCCCTCTC CCCTTGGCAA GTCAGTTGC CCTGATAGTA GCATGTTTCT GTTTCTGATG 480  
 35 TACCTTTTTT CTCTCTTCT TTGCATCAGC CAATCCCAG AATTCCCCCA GGCAATTTGT 540  
 AGAGGACCTT TTTGGGCTCC TATATGAGCC ATGTCTCAA AGCTTTTAAA CCTCCTTGCT 600  
 40 CTCTACAAAT ATTCAGTACA TGACCACTGT CATCTAGAA GGCTTCTGAA AAGAGGGGCA 660  
 AGAGCCACTC TGCGCCACAA AGGTGCGGT CCATCTTCTC TCCGAGGTTG TGAAAGTTT 720  
 CAAATTGTAC TAATAGGSTG GGGCCCTGAC TTGGCTGTGG GCTTTGGGAG GGGTAAGCTG 780  
 45 CTTCTAGAT CTCTCCAGT GAGGCATGGA GGTGTTCTG AATTTGTCT ACCTCACAGG 840  
 GATGTTGTGA GGCTTGAAAA GGTCAAAAA TGATGGCCCC TTGAGCTCTT TGTAAGAAAG 900  
 50 GTAGATGAAA TATCGGATGT AATCTGAAAA AAAGATAAAA TGTGACTTCC CCTGCTCTGT 960  
 GCACAGTCG GGCTGGATGC TCTGTGGCCT TTCTTGGGTC CTCATGCCAC CCCACAGCTC 1020  
 CCAGGAACCT TGAAGCCAAT CTGGGGGACT TTCAGATGTT TGACAAAGAG GTACCAGGCA 1080  
 55 AACTTCCTGC TACACATGCC CTGAATGAAT TGCTAAATTT CAAAGGAAAT GGACCCTGCT 1140  
 TTTAAGGATG TACAAAAGTA TGTCTGCATC GATGCTGTA CTGTAATTT CTAATTTATC 1200  
 60 ACTGTACAAA GAAACCCCT TGCTATTTAA TTTTGTATTA AAGGAAATA AAGTTTGT 1260

TGTTAAAAAA AAAA

1274

5

(2) INFORMATION FOR SEQ ID NO: 78:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1133 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

20

AGGATTTTTC	CTTGTTCAAC	CAAAATCTGA	GCATTCTTTC	TATGTTGAAA	ACACTGAAAA	60
ACTAATTTWA	GTTAATGAAC	TAGAAAGAAT	ATTGATTTTW	AAGAAACAGA	AAAATACTAC	120
TTATTTTCCT	TCTCAAATAA	CGTTTCTTTC	AAAACTTCT	GGCTGAAGTA	TAACATGCTG	180
GTAGTTAACA	TAAATCTTGT	CTTCTCTTTC	TTCTTTATCT	TTCTTTGTTA	TTTAGATGCT	240
TGTATAAATG	TCTTTTGTTC	TTATTAAGTG	CCTAATTGAC	AGAGCTTAAT	TTGAAGAAGT	300
GCCCTAATTT	ATTGACCACT	TAAGAATTGC	CTTTATTGGG	GTATTTTATT	TGTTCTGCG	360
TCTTTTGTAT	GTGTTTCAGT	CTACTCATCC	CTGTGAGTAT	GTGTGGGGGA	CAGCTGATAG	420
AAGGGAGGAG	AGTGTGTCTA	TGCTCAGGAT	TGCCCTTTAG	CCTCTCAGCC	AGAGATCCAC	480
AGGGAGCAAC	AAGGACAGTT	TCACATGCTT	AGACTTTCTT	GGAAGAAACA	GTGAGGAGGA	540
GTAAGTCGTG	AGTAGTGTCA	AGCTGCATGT	AGAATTGTCC	TAAGGCAGTT	GACCCACCT	600
TCCAACATGT	TTTCACTTTA	TTTGCCCTTC	CCTACATTTG	GGTTAGGTTT	CATTTGGATT	660
TGCAGCAATA	ATGACTTTAT	TTCTCTCTTG	GTGAGGATTT	GGCACATAAA	ATCCTTTTAT	720
TATAGAACTA	GCTATTTTAG	TTACATAGTA	ATGTAACATA	TGGAGAGATT	TATAGAGAAT	780
TTTGKTTTTG	CTGTATATA	TGTCCATTTT	GGAGACAGAT	ATGATAGAAC	TAGAAATTAA	840
GTTGCAATTC	TGCAAGTGCC	ATTGGAATGA	ACTTCAAGTA	TCTTCTTAAT	TATTAAATTT	900
TCTGATGAAG	GCATTGTAAC	AAATATATAG	TATTATTAAT	TCTAATTAAT	ATTGGAAT	960
ATTAATAAAT	AGGTATTTTA	TTTACTGTAA	AAAGTCAAAC	TTCATTATGT	AGATAAATCT	1020
TATTTCTTTC	ATTCTTTCCC	CTGTTTACAT	CCTTTTACAT	AAGCTTAGTC	ACCAATTAAA	1080
GCTTTCTCTAT	CAAAAAAAA	AAAAAAA	ACTCGAGACT	AGTTCTCTCT	CCT	1133

55

(2) INFORMATION FOR SEQ ID NO: 79:

60

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 661 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

GAATTCGGCA CGAGGGGAAA AGGATGCTGA ACGAGAGCAG AAAGCCTCTT TCCTTTGCTT 60  
 10 CACGCCCTTC CAGTCTTTAT TTAAACTCG GGTTCGCTTT CTGTGGTCCG AGCAACCTTT 120  
 ACTCCACCTG CACTGCTGCT COTGGGGGCT CCCCAGGCCT CCCTCTGCCT TTCTACCCAG 180  
 15 TGGCTGACGG GATGCCTGTC TTGCCTGGAC GCACCACTGC TCTCCTGTCC CTCACCTTGG 240  
 CTTTGTCTGT GCCCTGCTCT GGGGTTGAAG CTGGCCCATG TGTCCTCCCG AGTCATGGCT 300  
 GCTCCTCCTG GGAGGCCTCT GTGTGCGTCA CGTCTTCCAC ACCTGGGGGC AGCTGGCGAG 360  
 20 CCCGTGCTCT GTTCCCTCG GCTGCTTGCC ACAGAGYTGC AGCCTGGGAY TCTCCGTGGA 420  
 CCCAGACTGG GCATTTTGCC AGGGGGGCGA TGGGAGGAGC AGGTGCTTTG CCTGGCGGCT 480  
 25 GTGTCTGCAT TTCTGGACGC CCCAGAGCAC AGAAGTTGCC GGCACCTTGA GGTCTTCCTC 540  
 GGCATGTGCC AGATTACATG AGTGACCGCT GGAATATGT TTTCTTTTTT GTAATGGAGG 600  
 CGTGTTCAC ATATAGTAAA GTCACCAAAA AAGTAAAAA AAAAAAAA AAAAACTCG 660  
 30 A 661

35

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1378 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

45 ATTGGGTACC GGGCCCCCCC TCGAAGTTT TTTTTTTTT TTTTAATGAA AGCTCTCAA 50  
 TAAGCGATTT TATTCCTATC CATGATTGCA GACATTTACA AAACCATAAC ATCTGAGTTC 120  
 50 ACCTTAAAAA ATAACCTATA TAAAGCAGTG ATATACACAG CACAAAATAG TTCAGGGAGG 180  
 GGGCAGGAGC AACTTGTAAT AATTAAATC TAAACGTGAA AAAAAGGATG GAATAAAAGT 240  
 CCTACTTAT TTCTACTTAA GATGTCATGT GATAATATTT TACAATGTCC TGTGGGTCAA 300  
 55 TGTATGTATG TGTATATGTC TGTATAAGAT ACACATATAC AGTACATTCT CTTTCCCA 360  
 CATATACATA CACACATAAT TATTTCAGT TCAGTTTAGG GCAATTCTAA TATGCCACTC 420  
 60 CGTACAGTTC TTGAATCAC ATTGGACCC GCTTCTTCA CAAAAGACGG GAGAGAGCAG 480

335

5  
 10  
 15  
 20  
 25  
 30

GAAATAAAAA GGTGGTTTG GTGTGACTGA GATTCCCTTG TTTAACTGTA CACTGTGATG 540  
 AATAATTTTC TTCGCTAGTA GTTCTGTGAA GGGCTGACTC ACTGTGGTTT TCATGAGGAG 600  
 ACTTGGTAAT GGATCACACG CTCATTGTCA TGCTAGGGGA GTAACCTCTCA CTCTGAAAAG 660  
 GATTTAAGAA ATTTCCCCC ATTTCCGCCAT CATCCCTTGG AGTGGCCGGT TGATTACTCA 720  
 GGCTCATATT ATTGGGAGAA TTCTTGGAAA TACTGTCCAT ATCTCCTGAG CCTAAAGAGC 780  
 CATTCATGTG ATGTGACTCC ATTCCCTCCTA ATCCACCCAT GGGACCATCT GACCCAGGRC 840  
 CCATTGGAAA ATTAGGTCTG TTAGGTCCAG GAGGTACTGC ATTCATTAAG GTATACATGT 900  
 TATCACCAGA GTTGGTTGAA TCTGCTGGAC TAGGCATGAT GGGTGTTCCT GGTGGCCCTC 960  
 CACCTCCTGG AGGACCTACA TAATCCCAG GAGATGCTGA GGAGTATGGT ATTGAATTGG 1020  
 CATTTGTTGG GTTTGGCCAA GGTCTACCAC CACCTGGACC CATGTTTCATT CCAGGCATTC 1080  
 CAGGGCCACC TAAAGCATTC AGTGGGGGTC TCATTGCACC TCCATAGTTC TGTGGTCTTA 1140  
 AGGGCACCAT TCCTCTTGGG GGAGTCATTC TCTGCATTGG CCCACCCATA TTTGGATGTC 1200  
 CTTGTTGTGG AGTTGGATCC ATTCCACTGG GGAGTAATGG CTGACTTCCT GGGACACCTC 1260  
 CAAGTGCCTG ATTAGGTATC CTCAATGGGG/GCCTTGGACC TCCAGGGTAC CGAGGTGACA 1320  
 TAAAAGGGTA ATCATGGAAG GCTTTTGCTT CACTTGAGTG TTCACATGTT TCACGTCT 1378

## (2) INFORMATION FOR SEQ ID NO: 81:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1440 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

45  
 50  
 55  
 60

ACTTTGTCCA AATGTGTCTG TCACATCTAG TCAGCTGNAG NAATTTAAAA TGAATTGCCA 60  
 AGTGAAGAGT CTGTGGATTA ATTGGCCGTT AATTAACAGG CTTTATCAAT GTGTCCTCAA 120  
 GGGAGAGGCC CAACCCTAAT TAAGGAGCTA AACTTCCTGA GTGAGGGGCT GTGAGGATGG 180  
 AGGTGGAGGA GGCATCTGGG GCGGGTGGTG GCCGGGCCAG CAGATGGGCG CTCCCTGGCT 240  
 GAGCTGCCCC CACCGCCAGT TCCCTCATTT CCACTCAGGA AGGCAGAGAA GGCAGAGTGA 300  
 TCTCCTCAAG GAAGAGCTTC CCCAGCCTTC GGGAGCAGCT GGCAGGGCGT CCGGGAATAA 360  
 GCCCTACACG CCGCCGCTG CCTCCAACTC ACTAACCTG CGCCTCTTGT CTTTCAGATT 420  
 CAACCGCTTC AACAGAAGCC ATCCCCAGCC CAGCTTAAAT TATAAAGATA GACAATAACT 480  
 CTGTTCCAAT CTGCGTGGTG CTTCTTTAGT AAATACTGTA CAGATTTTAC CATGGAGAAC 540

TTTT TTTT TTTA GTTTT TACCT TTTCTTAATT ACCCTTATTC CGAATGGACG AACACTTTCT 600  
 ACCACTGCTG ACCATTGTAA AATACCGTGT ATATAAATCC CATTGAAATA ATGCCCTGGA 660  
 5 ATAGAACATC TCAAATGCTG CTTAATTACA GACTCAGGTC GATTACTTGT ATTTTCATGTA 720  
 ATGTTCTCTCC AAGTTAGACA TCTGGTGCAA GACCAACCGG GAGACCATGG AATTOTCAAA 780  
 10 AGTACAAACT GACAGTGTGT ATATTTAATT TAAAGACTTA TTTAAAACT CACAAGCTCT 840  
 CACCTAGACT TTGGAGAGCA GTCTGTTTTC TGTAAATGTCT GATACTAGAA ACTAATTTGC 900  
 TTATTTTAGT TGTATTCAAG ATTTGAAGAT GTATTTTATA GACAAGTTCT GTTTTGAAC 960  
 15 TTTGTGGAAC TGTTCCAATC AATCAATTTC CCAGTTATGA TGAGTATTTA CATTATGAAT 1020  
 GTATAACCCA GACATGATTT GTAAAGCCGA CAGTATGTTT CTATTACACA ACACTTTTTG 1080  
 20 ATACAGCGTC TCTTGTCTTC ACTGATACTG GAGTCTCCGT TGTCTGCMNG GTCCCTTCGA 1140  
 GTTCTAGTT ACAGACACAA TCATACTGTG ATTTTATTTT TAATATGGAT ATGCTATCAA 1200  
 ACTGTGATAC ACTTATAATT CACTGGTCTT GCATCAGGAG ATGGAGTGGG GAAAACTGTA 1260  
 25 TTTAATACAG TTTGTATCTG AATAATCTGT ATGGTTTATA CAGTTTGTGT TGTTCAGAGA 1320  
 TGTTTAAAGT TTGATCTTTG TTTTCTAAA GATTAAAAA GCACTTGCCC CACTGTAAAT 1380  
 30 ATACAGCATG TAAATTTCT RTAGTATATA AATGGCAGCA AATCACA AAAA AAAAAAN 1440

35 (2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1381 base pairs  
 (B) TYPE: nucleic acid  
 40 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

45 CCGGGGCTGC AGGAATTCGK YACGAGGCCA GCAGTTGCTC CCAGTTCAGG AGGTGCTCCT 60  
 GTACCCCTGGC CACAGCCCAA TCCTGCCACT GCTGACATCT GGGGAGACTT TACCAAATCT 120  
 ACAGGATCAA CTTCCAGCCA GACCCAGCCA GGCACAGGCT GGGTCCAGTT CTGACCTGAG 180  
 50 CACGGTTTTT CCTCATGTGA CTTCTGGGAA GGCGCTCCCT CATCTGGGOC AAAGGAAGGA 240  
 GGACGAAGCC CTCCTCAGCT GGCCTGTGTT TGGGGCATGA ATCTCTCCTC TCTCCTTGT 300  
 55 CTGGCTCTGT TGACAAACCG GGCATGTTTG GCAGTAAATT GGCACCGTGT CACACTGTTT 360  
 CCTGGGATTC AAGTATGCAA CCAGAACACA GGAGAAGAAA AGCTCCAGGA TCCCTGTCCC 420  
 CATCTGTCCT CTTGATGTGA GAGAGACTCT GAGACTTCTT CCATCGCAAT GACCTGTATT 480  
 60



	AAACACRAGC CCCCCAAGCA AAAGAAGAGG TTGAGTTTGC TGCCAGGATT CAGATCAGCC	540
	CTTCCCAGGG TCTGCAGGTG TCACATGATC ACAGTTCAGC GGGAGGCTTT CCGTACCCAC	600
5	ACTGGCTGTA GCACTTCAGT CCATCTGCCC TCCAGAGGAG GGTTTCTTCC TGATTTTATG	660
	CAGGTTTACA GCGTGCAGCT TGAGCTACAA TCAGGAGGGA AATTGGAAGG ATTAGCAGCT	720
10	TTTAAAAATG TTAAATATT TTGCTTTGCT AATGTGCTGA TCCGCACTAA CTCATCTTTG	780
	CAAAAGGAAC TGCTCCCTCG GCGTGGCCCA GCTGGGGCCT CTGAAGGGAT TCCTCACTGT	840
	GGGCAGCTGC CCGAGCTTC AGGCAGCAGT GTTCATCTCT GGCCAGTTGT CTGGTTTCCA	900
15	TGTATTCTAG GCCAGGTAGG CAACACAGAG CCAAGGCGGG TGCTGGAAGC CAGACGGAAC	960
	AGTGTTCGGG CAGGAAGGTG GATGCTGTTG TCATGGAGCT GTGGGAGTTG GCACTCTGTC	1020
20	TGCTGGTGGC COTCTCGGCT CACATGTTCA CAGTGCAGCT CCGGCAGAC TTGGGTTTTC	1080
	TCTTTGGTGG TTCTAAAGT GCCTTATCTG CAAACAATT CTTTTCTCCT TCAGGAAGTC	1140
	TGAATGGCTA GAAGAAGGAG CTCAGTAAAC TAGAAGTCCA GGGTTGCTTG GTTTACTGGT	1200
25	TTATAAGAAA TCTGAAAGCA COTCTGACAT TCCTTTTATT AACTCACCTC TCAGTTGAAA	1260
	GATTTCTTCT TTGAAAGGTC AAGACCGTGA ACTGAAAAAA GTGTTGGCCT TTTTCGGGGA	1320
30	CCAGATTTTT AAGATAAAAT AAATATTTTT ACTTCTGTCA AAAAAAAAAA AAAAAAATNT	1380
	C	1381

35

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 1706 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

45

50

55

60

ACTGCACCAC TGCCAGGTC TCCCGGCTGG ATGAAGACGT GGTCCATGAG GAAGCTGGCT	60
AGCTCAGACT GGAAGTAGC TTCAGGAAAA AAGACAAGTG GCCTAAGGAA ATCAGGCCCC	120
CCAACATCA TCTGAGGGCT AAAGATGAGA AGTAGATCAC TTAATAAGAC AAAAGCCTGT	180
AGGGGGAAAA GAAAGGATGT TTAAAAGGAC AGAATGTTTC CCAAGGTAGA AATGACACTG	240
TCAATTTCTC CTGGAATGG GGGCAGGGAT ACTCGCCTTG TTGCTCCGAC TTGAGTCAGT	300
ACTCACCTGC TCCTGGATCT CAGTATCCAC ATCTGAGAGG CAACTCTGGC AGAGTTCACA	360
GAAGGCCACC ATTCTGTCCC TCAAACTCGA CAGCTGCTTC TGTGGGCACA GTGGCTTGAA	420
GGGGARGAAT GAAGACACAG ACTCCTCTGT TCCCATATC CCATCTAAGA CCCCACTCA	480

CCTGGGGAAG CATCTGATTT AGAAATGTGG GTTAGTGTCC AGAGAATGGA AAAATAGACA 540  
 AGAGTCAAGG CTGGCAGGAT AACCTGTAAC AACAAAGGGT TTGAAAAATG ACGTTTGGGT 600  
 5 TAGGAGAGGG AGAGACAGAT AGCCAGAAAC ACACCAGTGA AGAGGAGAGA AAATCAGTAA 660  
 AGGGAGAGCT AATTCCTTTT CCAGTGGAAA ATGAGTGATA TTCTGGACAT TCTTCAGAGG 720  
 10 CATCTACAGG AAGTAGAAAT GTCACCGCTC CCTAATTTAC TCTACGTCTT CTAGAATCCC 780  
 TCAATATTAT CCTTGGCTTC CAGGAAATCC AAGAAGACCC TGGAAGTAGA GTCCACCTTC 840  
 TAAGAGAGGA ATGTAAGAGG TGACCCCCAC CCACCTGATC TTCCTCGCTT TGTCCACTCC 900  
 15 ACGCACTGAG ACTTGACACA CCTAGTGGCC ACCTAGAACG TAGGTCCTTA AAATVTAGCC 960  
 CCCCAGCCCC CAACCCATCT CTAGCCTGTC CACTCACCTG GTGAGGAACV TYTCCTGTGT 1020  
 20 CCACAGCTTT CTGCAGGAGT TGGCAACATG GTCATAGAG CTCCCAGCGA GTCAGGTCAT 1080  
 GAGTCTTTTG GGGGAGAAAG GGGAAATGTTA TACTGGAAAA GAACAGAGGG AACCAACTCC 1140  
 ACAGACACCA GTAAAAACGG GATGGGAAG AGGAGGAAAG CCACTCACTT GTAGAAGGCA 1200  
 25 GAGAGGCGTT TCAGAGTGGC TGCCAGATTA TATACCTCAT CCTCATCTAG GAAGGACGAC 1260  
 TGAGAAGGAA AGAAGATCCA CAATAGCATT TCCCCAGAA CTCATCAGTC CACATCCCCC 1320  
 30 GTCTTGACGC CCCTCCCACC CTTGTTTGGG GTGTCCCATT GTCCAGCCCC AGCTCCTACC 1380  
 TGTAACAGCT CTTCAAGCTC CTGCTGGAAR CCGTCAGTCA GCAAATCTAC TAGCTGGCTG 1440  
 CGGGCAAAGT CCGCCCGGCT GAAGAAAGTG AATTCGGGAT TACAGAGCAG GTAAGAGCAT 1500  
 35 GCGCCCCAGC CTCAAGCACC GCTGGCTCTG CATGCTTCAC CACCACCTCC TGGAGTTGCT 1560  
 GCAGGAACAG CTCCAGGTGC TGAGAAGAAA AGGCAGAAGA TGGTGTGCTG TGGGATGGG 1620  
 40 AGGAGGACAC TCTTCTGGCG GGAAGTGGAA CGGGTTTAAA AGCATTAAAC TTCAAGGATA 1680  
 AGATGCCTAA RAAAAAAAAA AAAAAA 1706

45

(2) INFORMATION FOR SEQ ID NO: 84:

50

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 573 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

55

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

60

GAATTCGGCA CGAGCTTGGT AGCCTTAGAA CTGCATGAGC TGCTTTACCA CTGGGAAACA 60

CGAGCAGAGC CTAGCTTGAT TTTGTATGTG GTATCAGATC TAAGGTGGAT GGAATTCAGG 120

ACTTCCTGTC TACTCTTTGA TTTTGTMTTA TTTTLAGAAA TGTTTTATTT TGTMTTATTC 180  
 ATTTATTCAT CTTACAGAGAC ATGGTCTGGC TCTGTTGCC AGGATGGAGT GCATGGTGTG 240  
 5 ATCATAGGCC ACTGCAGTGT TGAGCTCCCG GGCTCAGCG ATCCTCCTGC CTCAGCT/CC 300  
 TTAGTAGCTG GGAATATAGG CACATGCCCT ACCATGCCTG GCTTTGTCTA CTTTTTGAAT 360  
 GATGTCYCAA ACTAGAAGGT CTATTAATTT AAAAATTAA GGATAGCATG CCATAATTAA 420  
 10 AAATAATAAC AGTGGGAAAA GGCACCTTCC AATGATTCAG ACATCAACTT GTGATTTAAA 480  
 AAAACGAAAA ATAAATAATA GGAAAAAAG GCGAAAAAGT TAAATAAAAA TAAAATTAAA 540  
 15 AAAAAAAAAA AAAAAGTCGA GGGGGGCCCC GTA 573

20 (2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 684 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

30 CTCTTTGGCT GTGTCTACCT CCTTCATCTG CTGCGCGAC ATAAGCACCG CCCTGCCCT 60  
 AGGCTCCAGC CGTCCCGCAC CAGCCCCCAG GCACCGAGAG CACGAGCATG GGCACCAAGC 120  
 CAGGCTCCC AGGCTGCTCT YCAGTCCCT TATGCCACTA TCAACACCAG CTGCTGCCCA 180  
 35 GCTACTTTGG ACACAGCTCA CCCCCATGGG GGGCCGTCCT GGTGGGGGTC ACTCCCCACC 240  
 CACGCTGCAC ACCGGCCCCA GGGCCCTGCC GCCTGGGCT CCACACCCAT CCCTGCACGT 300  
 40 GGCAGCTTTG TCTCTGTTGA GAATGGACTC TACGCTCAGG CAGGGGAGAG GCCTCCTCAC 360  
 ACTGGTCCCG GCCTCACTCT TTTCCCTGAC CCTCGGGGGC CCAGGGGCAT GGAAGGACCC 420  
 TTAGGAGTTC GATGAGAGAG ACCATGAGGC CACTGGGCTT TCCCCCTCCC AGGCTCCTG 480  
 45 GGTGTCATCC CTTACTTTA ATTCTTGGC CTCCAATAAG TGTCCCATAG GTGTCTGGCC 540  
 AGGCCACCT GCTGCGGATG TGCTCTGTGT GCGTGTGTGG GCACAGGTGT GAGTGTGTGA 600  
 50 GTGACAGTTA CCCCATTCA GTCATTTCCT GGTGCAACTA AGTCAGCAAC ACAGTTTCTC 660  
 TGAACAAAAA AAAAAAAAAA AAAC 684

55

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1036 base pairs

60

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

	TGGAGGCAGA TGCACAGGAG AAAGGTTCCC GTCCGCACCC TCTCAGACCT GAGGCTGAGC	60
	TTGCAGTGAG GGCTTCTCCT CGGCCCTCG CCCGCCCCA GAGCTGCCAT CCCTGCTGTT	120
10	ACAAGCCAGA GGAGCCCGGA TGTGAGGCC CAGATCACCT CCAGGGACTT GGGGTTCCCA	180
	TCTGAAATCC TTTATTTTTC TACCATGGGG TGGGCCCGG GCTGAGAAGG AAGAAGCACC	240
15	CTCTCCCCGG CCTCCTCTGT CTGCACCGT GGGGCTGTGA CTTACTCCTG CCTCCAGGGG	300
	CGGGCCGGGG CCCCCTCGGA CCTCTTAAGG CCCAAGGTGG GCCCCAGGAC CTATGGGCAG	360
	AGTGGAYTGC TCATGGCAGA TGTGTGGCAA TGTCTGGCTG WGTCTTCCG GCAMCTGGCT	420
20	YCCCTYTCCC GGGYTCCCT GCTGCATGGT GGATGTGCTC CTTCCTGGCC CGGTACATT	480
	GCCTCCTTGA GCCTTAGTCC AGGGGGTCAC TYCTCCACC CCACCTACCT CACAGGGTTG	540
25	TTGTGAGGGT GCACAGAGGA GCAAGTCCC TGAAGGCCCT CAGGCAGTAT ATAGGGGCCG	600
	CCCACCTTCA GCTGCCCTGG GATGGGAAGG ACCCAGCCCG ACCCCTGGGC ATAACACTGT	660
	GTTTGCAAAT GGAGATTCAG GTATTGGGGA TGCAGGTTGT GGGGAGCTGG CCTGGCAGAG	720
30	TAGGGGTAGT TGGCTTGGCC TTCTCTTTGG TGATCCCACC CCCAGCCATT TGCATTGCTG	780
	GCCCAGCGCC TGGCCTGGGG GCGGGGAGA GGCAGCAGAA GGGGCTGGGC AGGGGCGGTG	840
35	GAGGACTCAG GAACTGCCCG GGGAGAGTGG GTATGGCGGC TGAGCCAGGG GCCTCCTGT	900
	GTTTGACTTC CCGGGATGGG TCCTTGCTTC TCAGCTGTGT CCGACCCAC CATGTAATAA	960
	AACCCAAAGG AACAGCAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1020
40	CCCNCGGGGG GNCCCC	1036

45

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:

50

(A) LENGTH: 908 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

55

60

	TTAAACAAAT GGAATCATGC AATATGTGAC CTTTTCCTC TGGCTTATTT TATTTAGCAT	60
	AATGTTTTTG AGGTTTCATC AAGCTGTAGC ATGTATCAGC ACCTCATTTT TTTTCTGGC	120
	TGAATATTAT TCCATTATAT GGATTTACCA CAATTCATTT ACCTATTGAT CTTTGTTTT	180

341

5 TGCTGTCTGG CTATTGTGAA TAATGCTTCG ATAAACATTC ATATACAAGT TTCTATGTGG 240  
 CTTTATGTTT TCATTCTCTT TGGCTATCTA CATGGCAGTA GAATTCTAGG TCATAATATA 300  
 ATTTTATGTT TAACCTCTCA AAGAATTGCC AAAAGGTTTT TCATAGTGGC TGCATCATTT 360  
 ACATTCCAC CGGCAATGTA CAAGGATTC TATTTTCCA TATCCTTGCA CTTACCAACA 420  
 10 CTTCTTTTTC GTWATWATTT TGTTTTTTCA TTATTGCCAC CCTAGTGGAT GTGAAATGCC 480  
 ATCTTATGTT TTGATTTGC ATTTCTCTAA TGACAAATGA TATCATACTT TTTTATGTG 540  
 CTTACGGATC AAAGGTATTT CCTTGGAGAA ATGTCCCTTC AAGTCCTTTG CCATTTCAAA 600  
 15 ATTIGGTTAT TTGTCTTTTA TTATTCAGTT TTAAGAAATT CTGGCCAGGC GCAGTGGCTC 660  
 ACCTGTAATC MTAGCACTTT GGGAGGCCAA GGCGGCAGA TCACTTGAGK TCAGGACTTC 720  
 20 GAGACCAGCC TGGCCAACAT GGTGAAACCC CATCTTACTA AAAATACAA AATTAGCTGG 780  
 GCGTGGTGGC AGGTGCATGT AATCTATCT ACTCAGGAGG CTGAGGCAGG AGAATCCCTT 840  
 25 GAACCCAGGA GCGGAGGCT GCAGTGAGCC AAGATCACGC CATTGCACTC TAGCCTGGGT 900  
 GACACAGA 908

30

(2) INFORMATION FOR SEQ ID NO: 88:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 655 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

45 TGCACTGGTT CCTTCTCCCC AGCAAATACT GCCTTCTTGT TTTTCTCTGA TGTGGCAGGT 60  
 GACTACAAA TCCGCCTTGG TATCTTCAA ATGCATATAT ATTCCTTTCT TGTCACTCC 120  
 CTCTCTTCT AGATTAGAAA ACTGCCTCAT TTTCTGCTCA CTGGATGTGC AGTCCCAGCT 180  
 TGTCTTCCTC TCCTCCCCCC CTGTTGCAGG TGTCTTTTTT TTTTCTTTC TCTCCCCACT 240  
 50 GGGCAGCAA AGTTGTTCCA CAGTGGAAAW TTAGGCATCC TCAAGTTTCY TCCCAGCTTC 300  
 TGCTGTGTTT TCCTAGAGTA AATTGCCAAT TTCTGTTTTT ACAGGAAATC CTTTTTTAAA 360  
 AATGGAATCA GTGTGGTCCC CATCTACTCT GCAAAAATTG CATTTTTCTC TATTTTCAAA 420  
 55 TGAGATTTGT TCAAGTTTCA AAACCACGTG AAATAATAAA TGTATAGTAG TTTTCTTTTC 480  
 CTTGGGCATT GCTWGATATG TGAAATGGGT TTATGAAAAA TAATAAAATC ATAACGCTAT 540  
 60 TTGTTTGAAT TTCAATTTC TGGGAATTTT TCTCGCTAA ACTCTAAATG GTGATTARGC 600

342

AAAAAAAAA AAAAAAACT GRAGGGGGGC CCGGTACCA TTCGUCCAT AATGA

655

5

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 1102 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

15

TTTTTTTTT ACCATTAAA ATAAATGAA AGTGACCTC TGTTTATAAA AATCTTGTG 60

TGCATCTCTG CTTATTTCCT TAGAAGAGAT TCCAGAAGC GGTGAGTGAT TTCACGGCAG 120

20 CAGAGGGTTG GGACATATTA CGGGCGCGGA TCCCTCTTGG AGTGAGATGA CTCTCCGGAG 180

AGATTTAGTC GTCACCTCG CGTGTGAGGC TCGCTCACAC CCCAGGGATG TGTCTATCAA 240

GATGGAAGAT CTTTACACG CTCTTGATT TGTGTGCTT TTTTCTATT ACTAGTGAGA 300

25

AKGAACTTT TTATATGATT ATTATCCATC ATAATCCAAC ACAAATTACT GCTTCATGTT 360

CTTTTACTTT CCGTGAAGG TTTTAGTGCC TTTTAAAAAT TGCTATATAT TAAGCTTGTT 420

30

AATACTTCCA TGCTGATTT GTGGSCATCA RTTCCCCCG GNACAGGCTT GCACATTTTG 480

CCTTCACACG CTGGGTGGTT TTTCAATTC AMTCTATTT CTCGTCTTC TATCGTTTTA 540

TGTCAGACG GTTTTCTCCG TGTAGAAAGC AGTTTATGAA GATTTACTTT CGACAGTCTT 600

35

CTCTCTACTT TCTACAGTGA ATTCTCTGAT GTGTCTGGA GTTTGGGGGT GTGGGTAAGA 660

RCCCTCCTCT CACCCTATTC TCTATTACGA TCCACAGCCT CATGCTTTAT GARATTGGTG 720

40

GCCCCGARGC GGGGAGATTT GGGATCCCC CAAGCCAGAC TTTATCCCC TATCCCTGCC 780

TCTGGATCCC ACCTACAGGC CTGGGAATTC CCTGTGGGTA GGGGCCAATG GTCTCGCACT 840

CTCACCTGTA CCCCAGGGCT GGCACAGGAT GGTCAAGGAG AGAGGCTGCC CAAGCGCATC 900

45

CTCTGGTGT CCCCCTGACA CGCCTCCAAA GTGAGCAGGT AGGTTTCAAC AGCCCCACGT 960

TGCAGGTGGG AGATGAAGCT CAGGTTGGAG ACCAGTATCT CACAGTTCTC TTTGCATGGC 1020

50

CGGGTACTTG TTAGTCAACT GATCAAGTGA AAATTCTAGC CCCAGAGGCA GGAGAATCCG 1080

GAACAAAATT AAACCAGCCA GG 1102

55

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:

60

(A) LENGTH: 1533 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

	GGCAGGAGCC GNCACGGGCA GCGCCCCATA GCGCCAGGGA CCCCCTGGCA GCGGGAGCCG	60
10	CGGGTCGAGG TTATGGATCC AGCGGGCGGC CCGCGGGCG TGCTCCCGCG GCCCTGCCCG	120
	TGCTCTGGTC TGCTGAACCC GCGCGGCGGC AAGGGCAAGG CCTTGCAGCT CTTCGGAGT	180
	CACGTGCAGC CCGTTTTGGC TGAGGCTGAA ATCTCCTTCA CGCTGATGCT CACTGAGCGG	240
15	CGGAACCACG CCGCGGACCT GGTGCGGTCC GAGGAGCTGG GCGCCTGGGA CGCTCTGGTG	300
	GTCATGTYTG GAGACGGGCT GATGCACGAG GTGGTGAACG GGCTTCATGG AGCGGCCTGA	360
20	CTGGGAGACC GCCATCCAGA AGCCCCCTGTG TAGCCTCCCA GCAGGCTCTG GCAACGCSCT	420
	GGCAGCTTCC TTRAACCATT ATGCTGGCTA TRAGCAGGTC ACCAATGAAG ACCTCCTGAC	480
	CACTGACAG CTATTGCTGT GCGCGCGGCT GCTGTACACC ATGAACCTGC TGTCTCTGCA	540
25	CACGGCTTCG GGGCTGGGCC TCTTCTCTGT GCTCAGCCTG GCGTGGGGCT TCATTGCTGA	600
	TGTGGACCTA GAGAGTGAGA AGTATCGGCG TCTGGGGGAG ATGCGCTTCA CTCTGGGCAC	660
30	CTTCTCTCGT CTGGCAGCCC TCGGCACCTA CCGCGGCGCA CTGGCCTACC TCCCTGTAGG	720
	AAGAGTGGGT TCCAAGACAC CTGCCTCCCC CGTTGTGGTC CAGCAGGGCC CGGTAGATGC	780
	ACACCTTGTG CCACTGGAGG AGCCAGTCCC CTCTCACTGG ACAGTGGTGC CCGACGAGGA	840
35	CTTTGTGCTA GTCCTGGCAC TGCTGCACTC GCACCTGGGC AGTGAGATGT TTGCTGCACC	900
	CATGGGCGGC TGTGCAGCTG GCGTCATGCA TCTGTTCTAC GTGCGGGGCG GAGTGTCTCG	960
40	TGCCATGCTG CTGCGCCTCT TCGTGGCCAT GGAGAAGGGC AGGCATATGG AGTATGAATG	1020
	CCCCTACTTG GSTATATGTC CCGTGGTCCG CTTCGGCTTG GAGCCCAAGG ATGGGAAAGG	1080
	TGTGTTTGCA GTGGATGGGG AATTGATGGT TAGCGAGGCC GTGCAGGGCC AGGTGCACCC	1140
45	AACTACTTC TGGATGTTCA GCGGTTGCGT GGAGCCCCCG CCCAGCTGGA AGCCCCAGCA	1200
	GATGCCACCG CCAGAAGAGC CTTATGACC CCGGGCGGC GCTGTGCCTT AGTGTCTACT	1260
50	TGCAGGACCC TTCCTCCTTC COTAGGGCTG CAGGGCCTGT CCACAGCTCC TGTGGGGGTG	1320
	GAGGAGACTC CTCTGGAGAA GGGTGAGAAG GTGGAGGCTA TGCTTTGGGG GGACAGGCCA	1380
	GAATGAAGTC CTGGGTCAGG AGCCCAGCTG GCTGGGCCCC GCTGCCTATG TAAGGCCTTC	1440
55	TAGTTTGTTC TGAGACCCCC ACCCCACGAA CCAAATCCCA ATAAAGTGAC ATTCCAAAA	1500
	AAAAAAAAA AAAAAAAAAA ANCCCGNGGG GGG	1533

## (2) INFORMATION FOR SEQ ID NO: 91:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 575 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

ATCCTCTGGA ATCTAGGTGG AAGCCACGAA GCCTTCTTCA CACTTGCGTT CTGAGCATCT 60  
 GCAGACTTAA CCCCATGTGG CAATCACCAA GCCTTATGGC TTGTGTCTC CAGAACTGTG 120  
 GCCAGAGCTG TACCTGGGCC CCTTTGAGCT GAGGCTGAAG CCAGAGTCTG AAGGTCAGCA 180  
 GGGCAGTARG CCCCTGGGCC TGGCCCTGA AACCATCTT TTCTCTAAG CCTCTGGGCC 240  
 TTTGATGGGA RGGGCTGTCC TCAAGATTTT TGAATGCCT TTGGAGGTT TTTGCTTGT 300  
 CTTGGATATT GCCTTCCTTT TAGTTATGCT CATCTCTTA GCAAGTGAAT GTTTCACAAC 360  
 CTGCTTGGAT TCTTTCTCTA CCACAGARCC AGGCTGCAA TTTTACAAAC TTTTACACTC 420  
 TGTTTCCCTT TTAAATATAA ATTTCAATGT TAAGTCACTT CTTTGCTCCC ATATCTGATT 480  
 TAGGTTGCTG GAAGTAGCCA AGTCACCTCT TGAATGCTTT GCTGCTTAGA AATTTCTCT 540  
 ACTAGGTAGC CTGGGTCATC ACACCTAAGT TCAAA 575

## (2) INFORMATION FOR SEQ ID NO: 92:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 639 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

TCCTTTCATC TTAAGCACCA CCCGACAGGG CAGGTACTAT TACCATCTCC GTTTGACAGA 60  
 TNAGGAACCT GGCACAGGAA GCATTTAAGT GGATTCCCCA GGATCGCCCC ACTGTCAGGA 120  
 GCAGANICAG AATGGGCGTC AGCATCAGGC TCCCAATCCT GGCTTCTAAC TGCTGCGCTC 180  
 TGCCCTTCYC TCWCCCCACC TCCCCACTCC AGTGCCTTTG GTCATGCCAC TGCAGCTTTT 240  
 AGGCCAATAC TGGATTAGCC TCTTAGTGT CTGTGCCCTG CAGCCATTTT CCCAGGCAGC 300  
 AATTCCATGT GCCCTCACTG ATGTAGGTGG CTCTTGTCTC ATTTGTCACT TCCTATTGAA 360  
 TTGTTTATGC ATCTTGTCTA CACTCACAGC ACCCTCCCTC TCACACGTC TCCTTATAAA 420  
 AATGTCCCTC AGTGTCTGCT ATGAGCCAGG TGCAGACTTA AGTGACAGGG CTGCTACGGG 480



345

AAATAAAAAA TTAACAAGGA GCACCTGCCT CTTAATGCAC AGTAACAAAC TATGTTAAGT 540  
GTCAGGAAGG AAAGGTTAAG GATGCCAGGA AGGCTTTTAA TAAATAACCT GACTTAGATG 600  
5 GGCAGGTGGT GCTGARGATT AAGAACGTGT TCTTCTCGA 639

10 (2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 744 base pairs  
(B) TYPE: nucleic acid  
15 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

20 GAATTCGGCA CGAGACTGGC TGGAGTCTGG CTGCAGAGGG AAGACATCAG CAGGGAGGGA 60  
GCCAGGCGCT GTCACATCTT TCCTCTGGCC ATTGTCTTGG TCTTTGTAAG CCCAGAATCT 120  
25 CCCCTTCCCT GAAGGGAGGC CAGCACCCCA GGAGGGCAGC AGGTGTGCTG TGAGGGTTGG 180  
AGTAGTGTGA GAGGTCAGGG TACACTAGAA: TGGCCATGGA CACCATGTGG GGGTGCTCTG 240  
GGCTGGGCA CAGAACAGTG TCCTTCCTGC TGCTCTCTCC CTGCAGCTTC CCCCAGCCTT 300  
30 GTNGTTTATT TGGTTTGATA CCAATCAGCA GACCCTGCAA GGTGGAAGCT CCCAGGCTCT 360  
CAGTCCCACS ACTCTCATGT GCCAGTCACC CNTACTGTAA CTGCCCAATG ACTACTTCTT 420  
GCCCACTGCC AAGATAGAGC CAGTTTACCA AGACAGGGGA ATTGCAGTAG AGAAAGAGTT 480  
35 GAATATACAT AGAGCCAGCT AAATGGGAGA GTGGAGTTTT CTTATTACTT AAATCAGCCT 540  
CCCTATAAAT TCAGAGGTGA GAATTTTCA AGGACAGTTT GGTGGSCAGG CCTAGGGAAT 600  
40 GGATGCTGCT GATTGGCTAG GGATGCAATC ATAGGGGTGT AGAAAAGTWC CTTGTGCACT 660  
GAGTCCACTT TTGGTGAGAG CTACCAAGGA GCTGCTGGTC TGCTGGTCCC GGTAGAGCCA 720  
45 TCTGGTGTCA GGAATGCAAA AGTG 744

50 (2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 526 base pairs  
(B) TYPE: nucleic acid  
55 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

60 GCAGGGGAAT TCGGCCACGG ACGGTTTCA ACAGGCTCTG TGGGGTGAGG TGCARACACA 60

346

AAGCCCATAA GTGCTGGCCT GTTGGGACAA ATGAGAGAAA TCCCATAGGG TGGTGATGAC 120  
 AGCGCAVTC A GCCATCTAT TCCCTGGGGAA AATGAACTT GTGCTCCTAT CAAATGCTCA 180  
 5 GTTGTAAC TCGAAAAA TTTAGAAGA CTTCTGTCC AGCATCTGTG TTTATGTCTA 240  
 TAAATGTAG AAACTAAAG CACAGAGATG TTAATGTTT TGTCCAAGGT CCAACAGCTG 300  
 GTTAGCARGC TTGGTCTGGT GACCTTTCTA CTGAACCACA GTGCCGCTGG GGAAGTCCT 360  
 10 CAGCAGAGAT GGCTGCTGCT ATAGCTGGGG TATGGGCACT ATTAGTAGTT AACCAGTCAA 420  
 CCCAAGTTCC CATAGTCTAG GTTCTGCTTC AGCTGGAGGT TAGGGAAAA CACAAGAAAA 480  
 15 TCCCTTACCA CTCTACCACT GCTGGGGGAT GTACTAAGAG ATCCCC 525

20 (2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 425 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

30 GGCACAGGGC AGGAGAGACT TGGTCCATGG GGAGAAGCCT GCAGTATAGA TGGGACCTCC 60  
 AGGAGCCCAA GTAGCATAGA CCTGCTGAT CCGGGGCCAT TGAGCCAGAG GATTGGGCT 120  
 35 GAATGTCCCC AGAGACAAA GGAAGGTA GATCCTTTCC CTAAAGATG AAAGCCATCG 180  
 CCGGGGCTTG CTTATGCTC TCTCTCTGG TCCTCCACA TGTGTTTCT GAACATTTGT 240  
 TCTGGCATCA CAATCCCCGT CATCCTGTCA TCTGGCCCTT CCCACCTTTC CACCTTATCT 300  
 40 CTTGCAGTGT CTCCGGCTCG ACCTGGCACC TGGGTGAARG CTTGCTCTTG CTGGTGCCCA 360  
 TAGCCCCCAG TGTATGTTCT TGAMCTCCCC AGCCATATGG ARACCCACCT CAGGAGGGCC 420  
 45 CCTCGA 425

50 (2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 844 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

60 GGCACAGCCG CACGAGATAG GAAGCTTGGC AGGGCCAGCT CCCCCAGTGC GCATTGCCCT 60

5 GTRAACTCGAG CGCCTGGGAG TGGGGAGAGG CTTCGAAATG GAGCAGGGTG GTGGACCTCG 120  
 TCTTCTCCTG CTCATCCCAG GCCTCCTCCA TAACACCTAC CTAGCAGGGC CTGGGGACTT 130  
 10 CCGAGCCCCA GGAACAACCTG AGAATACTGA GTGCCAGGGT AGCCCTAGCC CCATTTTACA 240  
 CCTGGGCAAA GTGAGGTCAC TGGATTCAA CACTCAGATT TAAACCTCCT CTGTGTCTGC 300  
 AGCACCTGTA TATAACTGCC AGCCTCTGCT GCGCCTCTCC AAAAAGTCTC TGGCCTTGTC 360  
 TTTGGCACCT GTCTCTGTCC TCCCCATTCT CTGCTCCTCC TTTCTCCAAC TCAGANTCAC 420  
 CCTGTTAGTT CAGCAAATGT TCATCGAGCT CCATAATGTA GCAGGACAGG NCTGTCTAAC 480  
 15 AGATTCTCGN CTTGCAAGGG TGAGACAAGT ACTCTCCATC TTTCTCTCAT CTTACAGAT 540  
 GGTCTGCTCA ACRACCTTTC ACTGAATPST AAATAATPGA TACTGCATAA AACATTGATG 600  
 TTCTTTAAGG GTAGTCCAGC AAGGTGGCAA GTCTTATAAT GATAACTGCT CAGGATCTC 660  
 20 TCAGTGAAGC ATTTGGGGST GCTAGCTCTG CCTATGGGTG AGGTCAGCTA TCTCAGGCA 720  
 TCTACTTCCA CNTGCCCCCC CATGCCAGGC TCACCCTGAG CTGAGATGCC TGACCAGGTG 780  
 25 GCAGAAAGGA GCCACCTGGT TTATGCTTCG GGACCACAAA CTGCTCTATC CAGANGACAG 840  
 TTTT 844

## (2) INFORMATION FOR SEQ ID NO: 97:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1985 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

35 AGCCCTGCTG AAGTACAGGT TCTTCTATCA GTTCTGTGTG GGCAATGAAC GAGCAACAGC 60  
 45 AAAGGAGATC ACGGATGAAT ATGTGGAGAC GCTGAGCAAG ATTTACCTGT CTTACTACCG 120  
 CTCTTACCTG GGGCGGCTCA TGAAGGTGCA GTATGAGGAA GTGGCTGAGA AAGATGATCT 180  
 AATGGGTGTG GAAGATACAG CAAAGAAAGG ATTCTYCTCA AAGCCATCGC TCCGCAGCAG 240  
 50 GAACACCAAT TACACCCTAG GAACCCGCGG CTCTGTCTATC TCCCCACTG AACTTGAGGC 300  
 CCCCATCTTG GTGCCTCACA CAGCCGAGCG GNAGAGCAGA GGTATCCATT TGAGGCCCTC 360  
 TTCCGCAGCC AGCACTACGS CTTCTAGAC AATTCCTGCC GCGAATACCT TTTCATCTGT 420  
 55 GAATTTMTTG TTGTGTCTGG CCCAGYTCCA CACGACCTGT TCCATGCTGT CATGGGCGGT 480  
 ACACTCAGCA TGACCCTGAA ACACCTGGAT TCTTATCTAG CTGACTGCTA CGATGCCATT 540  
 60 GCTGTTTTTC TCTGTATCCA CATGTTCTC CGGTTCCSTA ACATTGCAGC AAAGAGGGAT 600

GTTCCTGCCC TGGACAGGTA CTGGGGAACA GGTGCTTGCC TTGCTATGGC CACGGTTTGA 660  
 ACTGATCCTG GAGATGAATG TTCAGAGCGT CCGAAGCACT GACCCCCAGC GCCTAGGGGG 720  
 5 GTTGGATACT CGCCCCCACT ATATCACACG CCGCTATGCA GAGTTCTCCT CCGCTCTTGT 780  
 CAGTATCAAC CAGACAAATC CTAATGAACG GACCATGCAA TTGCTGGGAC AGCTGCAGGT 840  
 10 GGAGCTGGAG AATTTTGTCC TCCGAGTGGC AGCTGAGTTC TCCTCAAGGA AGGAGCAGCT 900  
 TGTGTTTCTG ATCAACAACT ATGACATGAT GCTGGGTGTG CTGATGGACC GGGCTGCAGA 960  
 TGACAGCAAA GAGGTTGAGA GCTTCCAGCA GCTGCTCAAT GCTCGGACAC AGGAATTCAT 1020  
 15 TGAAGAGTTG CTGTCTCCCC CTTTGGGGG TTTAGTGGCA TTTGTGAACG AGGCTGAGGC 1080  
 TTTGATTGAG COTGGACAGG CTGAGCGACT TCGAGGGGAA GAAGCCCCGG TAACTCAGCT 1140  
 20 GATCCGTGGC TTTGGTAGTT CCTGGAAATC ATCAGTGGAA TCTCTGAGTC AGGATGTAAT 1200  
 GCGGAGTTTC ACCAACTTCA GAAATGGCAC CAGTATCATT CAGGGAGGCG TGACCCAGCT 1260  
 GATCCAGCTC TATCATCGCT TCCACCGGGT GCTGTCCAG CCGCAGCTCC GAGCCCTCCC 1320  
 25 TGCCCGGGCT GAGCTCATCA ACATTACCA CTTTATGGTG GAGCTCAAGA AGCATAAGCC 1380  
 CAACTTCTGA TGTGCCAGAA ACCGCCCTGA GATCTGCCGG TCATCTCCAT GGACTTCTGC 1440  
 30 ACCCCATTC ATACCCTTCT TCACCTGGGG TACCCCTTCC AGTTTTCCCC TTGCTTCCCA 1500  
 GGCCCTTGAC ATGGCTTACC TGCCTTCACT CCCAGCACCT TGCCCAACAG GATAAGCTGG 1560  
 ATCCCTTGG CTTTCTGAAT ATCCAGTGT CTTGAGTTT CCAAGACCA CTTCCCTGTG 1620  
 35 GGCTTCCAAA ATGGCCTTTA TCATTTCTCC AGTCTGTCAC CTTCCCTTCC TGCTCCATA 1680  
 CACCCAAGGC TTGTTTCTTC CCCTGTAAAA ACCACTGCCT CAATCTCTGG TTCACTCAAC 1740  
 40 TAGTCACCAT GTCTGAGGC ATGAAGCCTC CTCAGCTCTT GGAATTGCTG GCAAGGGGTG 1800  
 ACTGCCTCTG AGTCATTGTG TTTTCAAAG TGATTCTTT TCTGTAGCTT TTTGACCTAA 1860  
 GATCTCAGCA ATTTGAACAC TAACCTCTCC CTTCTGGCT CAAGAATTAC TCCGAAGTCA 1920  
 45 GTCTGCAGAA AATAAATATT TAGTATGACA TGAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1980  
 AAAAA 1985  
 50

(2) INFORMATION FOR SEQ ID NO: 98:

- 55 (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1416 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- 60

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

	ATATGAAGGG AAAGAATTTG ATTATGTTTT CTCAATTGAT GTCAATGAAG GTGGACCATC	60
5	ATATAAATTG CCATATAATA CCAGTGATGA CCCTTGTTA ACTGCATACA ACTTCTTACA	120
	GAAGAATGAT TTGAATCCTA TGTTCCTGGA TCAAGTAGCT AAATTTATTA TTGATACAC	180
10	AAAAGGTCAA ATGTTGGGAC TTGGGAATCC CAGCTTTTCA GATCCATTTA CAGGTGGTGG	240
	TCGGTATGTT CCGGGCTCTT CCGGATCTTC TAACACACTA CCCACAGCAG ATCCTTTTAC	300
	AGGTGCTGGT CGTTATGTAC CAGGTTCTGC AAGTATGGGA ACTACCATGG CCGGAGTTGA	360
15	TCCATTTACA GGGAAATAGTG CCTACCGATC AGCTGCATCT AAAACAATGA ATATTTATTT	420
	CCCTAAAAAA GAGGCTGTCA CATTTGACCA AGCAAACCTT ACACAAATAT TAGGTAAACT	480
20	GAAGGAACCTT AATGGAAGTG CACCTGAAGA GAAGAAGTTA ACTGAGGATG ACTTGATACT	540
	TCTTGAGAAG ATACTGTCTC TAATATGTAA TAGTTCTTCA GAAAAACCCA CAGTCCAGCA	600
	ACTTCAGATT TTGTGGAAAG CTATTAAC TGCTGAAGAT ATTGTCTTTC CTGCACTTGA	660
25	CATTCTTCGG TTGTCAATTA AACACCCAG TGTGAATCAG AACTTCTGCA ATGAAAAGGA	720
	AGGGGCTCAG TTCAGCAGTC ATCTTATCAA TCTTCTGAAC CCTAAAGGAA AGCCAGCAAA	780
30	CCAGCTGCTT GCTCTCAGGA CTTTTTGCAA TTGTTTTGTT GGCCAGGCAG GACAAAAACT	840
	CATGATGTCC CAGACGGAAT CACTGATGTC CCATGCAATA GAACTGAAAT CAGGGAGCAA	900
	TAAGAACATT CACATTGCTC TGGCTACATT GGCCCTGAAC TATCTGTMTT GTTTTCATAA	960
35	AGACCATAAC ATTGAAGGGA AAGCCCAATG TTTGTCATA ATTAGCACAA TCTTGAAGT	1020
	AGTACAAGAC CTAGAAGCCA CTTTATAGCT TCTTGTGGCT CTTGGAACAC TTATCAGTGA	1080
40	TGATTCAAAT GCTGTACAAT TAGCCAAGTC TTTAGGTGTT GATTCTCAA TAAAAAGTA	1140
	TTCTCAGTA TCAGAACCAG CTAAAGTAAG TGAATGCTGT AGATTTATCC TAAATTTGCT	1200
	GTAGCAGTGG GGAAGAGGGA CGGATATTTT TAATTGATTA GTGTTTTTTT CCTCACATTT	1260
45	GACATGACTG ATAACAGATA ATTAAAAAAA GAGAATACGG TGGATTAAGT AAAATTTTAC	1320
	ATCTTGTAAG GTGGTGGGGA GGGGAAACAG AAATAAAATT TTTGCACTGC TGAAAAAAA	1380
50	AAAAAAAAAA AAAAGGAAAC TCGAGGGGGG GCGCGG	1416

## (2) INFORMATION FOR SEQ ID NO: 99:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1935 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

5	NTCTACCCTA ATCAAGATGG GGACATACTT CGCGACCAGG TTCTTCATGA ACATATCCAG	60
	AGATTGTCTA AAGTAGTGAC TGCAATCAC AGAGCTCTTC AGATACCAGA GGTTTATCTT	120
	CGAGAAGCAC CATGGCCATC TGCACAATCA GAAATCAGGA CAATAACTGC TTATAAAACC	180
10	CCCCGGGACA AAGTGCACTG CATCCTGAGA ATGTGCTCTA CGATTATGAA CCTCCTGAGC	240
	CTGGCCAATG AGGACTCTGT CCCTGGAGCG GATGACTTTG TTCCTGTGTT GGTGTTTGTG	300
	TTGATAAAGG CAAATCCACC CTGTTTGCTG TCTACTGTGC AGTATATCAG TAGCTTTTAT	360
15	GCTAGCTGTC TGTCTGGAGA GGAGTCCTAT TGGTGGATGC AGTTCACAGC AGCAGTAGAA	420
	TTCAATAAAA CCATCGATGA CCGAAAGTGA CCAAGACCAA GCGCCACCAA GGCAGCAGAC	480
20	TGTTAATCAG ACAAACAGAT CTCTGAGAAG GTGCATCAGC TGCTTTGAAG GCTGAAGATT	540
	GTTTTGTATG ATACTGCACA GCATCAGGCA TTATAAGCA GATCTTTACT AAACAGGTTA	600
	ATGAGCTAAC AAGCAGGTTT TCTCGTCTTT GGGCTCTTTT CTTTCTGAGT TGCATATTCT	660
25	ATTTTCTTGT CCCCAAGTAG AGACTAGTAG TACAAAAAGG GACCACATTT TTCAAGTATT	720
	TCTAAGTATA AAAAACAAAA CAAAAATCTC TTAGGAAATG TCTAGACCTC CATCTTGGA	780
30	TTCCCTTTCT TTCTTTTAT TTAAAAAAG AACAGTACCC CTCTTTTAAG ATGCTGTCTT	840
	ACATTAAATGA GCATCTAATG GAAAGAAGGT ATGAGTTGCA CTGAGGATTA GAATAGTGGT	900
	GCGTTAGTGG CATTATCTAT AAATACACTC ACCTAAATG AAAGCTAAGA AGGAAATGTA	960
35	AATATAATAT ATATTTATAT TTGATGTAAT ATGGACATCT GCAGATTCTA ATAAACAAGG	1020
	ACTATTGCTG ATAGTAGGCT GTGACATACT GTCTTGTAAG ATGGTTTCCT TGACAAAATT	1080
40	TAAGCTGAGC TTAAGGCAA AAAAACAAAA AGTACACAGA AATATTTATT AAAATGTAAT	1140
	ACAGTTTATT GAACTTTCTA GGTATGGAGT TTGATGGACA GGGCTGCCTT TAATGAGTGT	1200
	GAAGGTCAC T AAGTCACCTA GACATCTCAC CGTGGAAATT TGTGAGCCTG CATTAGGAGA	1260
45	TAGACTGATT ACCATACATG ACATAAAAAG GAACAGTGGG TAGCTCATAC TTTATGGTGG	1320
	TTCTTCTCCT CCGAAATAAT ATACTGCAGA AATCCCAGAC AGAGCTCCTT ACAAACCTTT	1380
50	AATTGTAATA TATTTTGTAT GATTATTTCAC ATTGAATGCA CAGACCAAGA ATTCAGTGAA	1440
	TGTCAFTTTT TAAAAACTA ATTGTAATG TCTGCTCTAG TGATACAAGT TTTACTAGTG	1500
	ATAAACTATT TTAATCAACC ATACTATTCT TATGGAAAAA AATATCTATT TTGGCAGGTT	1560
55	TCTGTGCCTT TATTTCCCTC TTCTGAAAAA AAGTCTGTGT TTTCATAGTT TGGTTTGCAT	1620
	TGTATATCAA TAATTAATCA GGAATGGGTT TTGGTGCCTG AAAAATTGGC CATGGAGCCA	1680
60	CACCAAAGCT TCAAGCACAA GTCTTGTAACA TGGGCCATCA CTGTCTGGTT TCACTTCGTG	1740

TGTTTCCTAA ACACATTTAG CTGCTTTTTT AACAACTCA GCCCCATACT TGAGTCCCTT 1300  
 GTTGTTGGGA GCATTTCCAG GCATCTTTTA AGGGAAGTGT GACAAACAGC CTCGGGCAGA 1860  
 TGAACACGGA GGCTCTCTGT TGTCTGTCTC TGAGATCTTT GTGTCTGGGA ATGCCTAAAG 1920  
 NTTTTGNTTT TTTT 1935

(2) INFORMATION FOR SEQ ID NO: 100:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 599 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

GAATTCGGCA CGAGCGTCCA CGCAGCCGCC GCGCGGCCAG CACCCAGGGC CCTGCATGCC 60  
 AGGTCGTTGG AGGTGGCAGC GAGACATGCA CCCGGCCCCG AAGTCCTCA GCCTCCTCTT 120  
 CCTCATCCTG ATGGGCACTG AACTCACTCA AGACTCCGCT GCGCCCGACT CCCTGCTGAG 180  
 AAGTTCAAAG GGCAGCACGA GGGGGTCTTT GGCTGCTATT GTCATCTGGA GGGGAAGAG 240  
 TGAGAGCCCG ATAGCCAAGA CCCAGGCAT TTTCAGAGGT GCGGGGACCT TAGTCCTACC 300  
 CCCAACACAC ACCCCTGAGT GGCTCATCCT CCCTTGGGC ATAACGCTGC CTTGGGGGC 360  
 TCCAGAAACA GCGGTGGGG ATTGTGCCGC TGAGACCTGG AAGGGCAGCC AGCGTGCCGG 420  
 CCAGCTGTGT GCATGTCTGG CTTAATATGC AGGCTTGGG GGGCTGTGGC CACATGCCCG 480  
 GCAGGAGGTG AGTGAGGAGC CCTGTGGCGT GCTGGTGTGG GGATCGTGGG CATTTCAAAC 540  
 GGGCTTGTG TACCCTGAAC AATGTATCAA TAGAGAAAAA AAAAAAAAAA AAAACTCGA 599

(2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 784 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

GAATTCGGCA CAGAAAAAAG AGAGAGACTG GGTCTTACTG TGTGCCCCAG ACTTGTCTTG 60  
 AACTCCTGCC TCAGCCTCTC AAGTACTTGG GATTATAGGC CAAGAAGCCA CCATGCCTAG 120  
 CTCTCTCTG TCATGTATCC AGACTAATAC TCTGGGTCA GCTCATTTT TTCTCTTCT 180

	CACTTTGCAC ATCCACTTGT CACCAAATCK RGTTCATTCT GCATCCTAAG TAAGTCCTTT	240
5	GATTCCTCCA GTTGTTCATT AGTAATGTCT CAARTGTAAT TTTTCTAGT AGTTTTCAGC	300
	CTGTCTTTCC KGCCTTCAGT CTTAACTTCT CCAGTACATA KGCCACATTG TTGTCAGCAK	360
	GATCAWATTT TATTTAAAAA TACTTTACAW AKGTTTATKG CCAAATATTA GRAAATACAG	420
10	ATTCATGGAA AGAAAAATCA CTGTCCCAAG GAGGTCACCT GCATGGTGAG GTTAAGGGGT	480
	GATTTTAATT TTTAAAAATG TATATTTTTT CCTGTGTAGA GTAGTAACAC CCTTGAAAAC	540
15	ACAWTCCCTT GTAAAGTCTC TAATTCTGTA CTCCGCATCT AGSTGRTCTC TTCTTTCTCA	600
	GATATTTTAC AATTTCAATT ATCACCACCT TTCTCTAGCC TTTACCCGTC TCTTCAATAT	660
	TWACATATGC AGAAGTTTCT CCTAACAAAC ACCTGCCTCT GCCTCAGTTC TGCTACCACC	720
20	CTGTGCTTT CTTTCCCTTC ACAATCAAAT TTAAGAGTGT CAAAAA AAAA AAAAAC	780
	TCGA	784

25

(2) INFORMATION FOR SEQ ID NO: 102:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1035 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

	AGAGGCCTGG CTGGTTGCC CTATCTCCGT CTCCGCCACC CACTTAGCGT TTAGGCATC	60
40	AATTACCAGC AGTTTCTCCG CCACTATCTG GAAAATTACC CGATTGCTCC CGGCAGAATA	120
	CAAGAGCTTG AAGAAGCCCG CAGTTCCGTG GAAGCCTGCA GAGCAAGGGA AGCAGCGTTT	180
	GATGCCGAAT ATCAGCGAAA TCCTCACAGG GTGGACCTCG ATATTTTAAC CTTTACGATA	240
45	GCTCTGACTG CCTCTGAAGT TATCAACCCT CTGATAGAAG AACTTGCTTG CGATAAGTTT	300
	ATCAATAGAG AATAGTTAGG TGGTGACACT ACTTCAAGAG AACCTCTCCA TTCCAGTCAT	360
50	ACCAATCCTG CAACTTGATT TTCAGAAGTC AAGAGTATAT CGCGATAAGA CAGTGCACAG	420
	GTGGAGGGGA AAAAAAGGGG GAGCGGGAAG CTTATCTTGA AAAAGCATCA CAGAAGTAGA	480
	AAAAAATGTC GAAAGCATT AACTGTAAAC GTTCTTTGAG TTTGTGATTG ATCCACATTT	540
55	TTCCCCCTGC ATTATGGAAA ATGTCTCTCA GCATTGCTTT ATTACAAAGT AAAGGATGGT	600
	TTTATAAAT TGAGACTGAT GAAACATCAA TACTAGAGCC CATGAGGATG AAAGAAATTA	660
60	TCAAATAGTG CTGAACAGAA TAAGATGTTA ACGCTGAGTT ATTAGGACTG GAAGGCTATG	720



353

AAAAGAACTT GAAATGTGCG GAATATGTGC TCTCTTCATG TCATATTCAA TAGAAGTTTC 780  
 TAGTTTAAGA TTGATTTTGT GTTTTCTTAG GCATTTCAAG TGACAAGCAA AGTAAATGTA 840  
 5 TATATTATGT GATAAATCAT GTTTTCAAGA ACGTCAAATT TCTGGACTTT TTTCTTTCAA 900  
 TTTTAAATTT TTAAAGTTTT TTTGGTATTA AAAAATCTAT TCACAAGCCA AAAAATWTWT 960  
 WAAATWTWCM GCGAAAAGCC AAAAAAAAAA AAAAMMAGGG GGGGCCGGGC CCCATCCCCC 1020  
 10 CAAGGGGGTC CNGMT 1035

15

(2) INFORMATION FOR SEQ ID NO: 103:

(1) SEQUENCE CHARACTERISTICS:

20

- (A) LENGTH: 2218 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

30

35

40

45

50

55

60

AGGTATTAGG CCCTTTTGTG GGAGCCCCAT GTTTTGTGTTT TCTGAGTTGG TGGGGAGGGA 60  
 SGGAGGGGGA GGGCTGAATT GTTTTGCAGA GGAAGATGCC ATCTGTGCTT TAAATTTCTC 120  
 ATTACTGGGT TAGAAAACAA AGAGGGAKTG CCCTGCACAT TTTCTTTTGT GCTTTTAAAT 180  
 GTTTCTTAAG TTGGAACAGG TTTCCTCGGG CCTGTTTGA CTGATTGCTG GAGTGCATTT 240  
 GATAGTTAAA AATTACTAAT TGGTTTTATT TCCCTTCACA CTCTGCCTCC CCACTTCTCC 300  
 CCCCCTTACT GAAAAATAAC CATTTTAGTG TCAGGCTAGA AATTGAATTG CTGAGTTTTC 360  
 TGTATCCTTT AAATTAAAA CCACAAGTGT TTATTGTAGT GGTAAACTG TAGCATCTCA 420  
 GCATCTGGGT GGAAGCTGCC TATATTCTT CCCAGTTTAA CTGGGGACCA TCTGTGAAAT 480  
 TAATTTTCCA TCCAGACAGC TGCTGTGAGC AAATGAACAT AAATGCTGGC TGGAAATTTA 540  
 CTAACCAATT TTTATATTGA CCTGCAGTGT AAAAAGCACA TTTAATTATA AACCAATAT 600  
 TCAAAATGGG CAAATTTTAT TTTCAAATGC AGTGTAGAGC TAGATTAAAA GCAACTCTTT 660  
 GCCACCTACT CTGCCCTTTT GGCAAAGTTA CCTTGAACAA AGAATCTTAA GGGTTTATTA 720  
 AGAACTCTTT ATTTTCTTCA TACCCTGTTT TCTGCAGTGC TTTCTAACAG CTTCTGGGTC 780  
 CAGATTTTCT TCGGCATCTT TTTGCACTCA GCTTATTACA GGTAGGTAGT GCTTAAGAAA 840  
 AGTCATGGAG GACTAAAGCC TAAGTCCTTT TCACTTTTCC TCCATCTGAA GGTAGGTGAG 900  
 TTCATCCTCT TCATAGTAAT GCTGTTTTAC CAAGACTTTA TAGCAGATGG ACCCAGAAAG 960  
 AATTTTCTGC TATTGTGTTT ACTACACAG GATAGGGACA TCAGACAGCC CCAGAAACCC 1020  
 CTTCCAGATC TGATATGGGA CTATTAAATT TTATGCTGTT AATTGGTATT CATTCACAAT 1080

GCAGTTGAAG GGGGAAGGCT CCACTGCATT CTTTGGCTAA GGCCTGAATG CTTGCTCATC 1140  
TGTAAGATCT ATACTCGAGG TTTTGTTC CTTTAAAT TCTTAGGGA GAGAGGGATG 1200  
5 GTTCTGAGG GGTCTGAAA GTATGATTCA ATGTGCAACA TACAGGTAGG TCTTCAGCAT 1260  
AAGCTGAAAT ATATGCATGT AAAAAGTTTG ACATCTTTT TTTAATTTT CCACTTCTTT 1320  
10 CTTAACTTTA CTTCTCTTTT TGTCCCCCCC CCATCTTACA GAAGTTGAGG CCAAGGGAGA 1380  
ATGGTAGGCA CAGAAGAAAC ATGGCAAAC GCTCTGTGCT TTCAAACCAA AGTGTCCCC 1440  
CCAAACCCAA ATTTGTCTAA GCACTGGCCA GTCTGTGTG GGCATTGTTT TCTACAACCA 1500  
15 AATTCTGGGT TTTTTCTTC TTTCTTAAA CATAGAGGTA CCACCACAAG GGATGCCCTA 1560  
CTCTCTCGCA GCTCTGAAA GCATCTGTTT GAGCGAAAGG TCTCTGGGCA AGCAAGTGGT 1620  
20 TATTTGGATT GCTGTCTCC CTTTTTCCAC CTGGGACATT GYAATCATAA AATAACAGTA 1680  
AATTCCAAAC CTCAAAACT ATTATGGCCT GAGCACAGCT GAAATCTAGC AGAGTTTAC 1740  
TCTTCTGCCT CCATGTCTGT CACTTATAAT TCAGGTTCTG CTGTTGGCTT CAGAACATGA 1800  
25 GCAGAAGAAT CGTTTTATGC TAGTTATG GATCATGGTT GAAACTCAAC TTAGGGAAAG 1860  
GGTTCCAATG TATTAAGCAA TGGGCTGCTT CTCCTCAATC CTCCTAACA ATTGCTGTG 1920  
30 TGGACTTCTC ATCTAAAGG TTAGTGGCTT TGGCTGGGA TCAGTGCTCT CTATTGATGT 1980  
TCTTGCTGGT CTCCAGACAC ATTCTGTTG CATTAAAGCT TGAAAGACTT GTAGATGTGT 2040  
GATGTTCAAG CACAGGATGC TGAAGCTAT GTTACTATTC TTAGTTTGTG AATTGTCCTT 2100  
35 TTGATACCAT CATCTGTTT TCTTTTGTG GGTATAAATA AAAACACTGT TGACATATAA 2160  
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAA 2218

40

## (2) INFORMATION FOR SEQ ID NO: 104:

45

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1351 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

CTTCACAGAC TGACAGAATG GTTTGTTTT GTTTGTTTT GTTTGTTTT GTTTTGAGA 60  
55 TGGACTCTAG CTCTGTACAC CAGGCTGGAG TGCAGTGGTG CGATCTCGGC TCACTGCAAG 120  
CTCGGCTCC CGGTTTCTCA CCATCTCTCT GCTCAGCCT CCGAGTAGC TGGGACTACA 180  
60 GGGCCCCACC ACCACGCCCG GCTAATTTTT TGTATTTTT ACTAGAGACG GGGTTTCACC 240

355

	ATGTTAGCCA GGMGTGCTCT GATCTCCTGA CCTCGTGATC GSCCCGC/TC GGCTCCCAA	300
	AGTGCTGGGA TTACAGGGGT GAGCCACCGT GCCTGCCCCA GAATGGTTTT TAAAGCCACA	360
5	GTTGAGARGC CACCCATTGC CCGGCGCCTG GACAGTGATC ATCTTGTTCA TCTTGTTGAG	420
	TCCTTTCTTG TGTGATTGGA ATTATTGATC CCCTTTGAAA GATGAGAAGG TTGAGATGCA	480
10	AAGAGTCTAC CTTTCCAAGT TCTCACTGCT GGAAGARCT AGAAGCACAG TTCAAAGTTC	540
	TGNNTTCTGG ACTCTGCAGT CCAGGTATCC CTMTTCCAC TTGCCTACCC TCAATGCCAC	600
	ACTGTTTTTG AAGTGGCCCA TAACTTGAAG GRAAAGTTA AAGACAGTTC AATTTAATCA	660
15	TCAGRATGCA TTCTTTTTTT TTTCGGARAC GGAATTTTAC TCTTGCTGCC CAGCTGGAG	720
	TGCAATGGTG CAATGATCTC GGCTCACTGC AACCTATGCC TCCTGGGTTT AAGNGATTAT	780
20	CCAGCCTCAG CCTCCCGAGT AGCTGGGATT ATGGGGCCCC ACCACCATGC CCAGCTAATT	840
	TTTGTATTTT TTTTCTTACT AGAGATGGGG TTTCGCCAGG TTGGCCAGGC TGKTCCTGTG	900
	AATCCTGGC YTCAGGTGAT YTGCCACAT CATCTTCAA AAGTGCTGGG ATTACAGGCA	960
25	TGAGCCACTG CGCTGGCYT CAGAATGCAT TCTTACACAT CTATCCTAGA CATTTATAAG	1020
	CACTCTAATG GATAACAATC CAAGAATAAA TGATTGTAAA AGATGATGCC GAAGAGTTGA	1080
30	TGTCATCTT TTTTCTCTAA GAAAAAAGT CCGCGAGTAT TAAATATTTA GATCAATGTT	1140
	TATAAAATGA TTAATTTGTA TATCTCATTA TTCCTATTTT GGAATAAAAA CTGACCTTCT	1200
	TTAATCATAT ACTTGTCTTT GTAAATAGC AGCTTTTGTG TCATTCTCCC CACTTTATTA	1260
35	GTTRATTTAA ATTGGAAAA ACCCTCAAAC TAATATCTT GTCTGTTCCA GTCTTATAAA	1320
	TAAACTTAT AATGCATGTA AAAAAA A	1351

40

(2) INFORMATION FOR SEQ ID NO: 105:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2066 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

45	GGCAGGAGGC GCGGAGGGC CACAATCACA GCTCCGGGCA TTGGGGGAC CCGAGCCGCC	60
	TGCGCGGGGG GAATCCGTGC GGGCGCCTTC CGTCCCGGTC CCATCCTCGC CCGGCTCCAG	120
55	CACCTCTGAA GTTTTGCAGC GCCCAGAAAG GAGGCGAGCA AGGAGGGAGT GTGTGAGAGG	180
	AGGGAGCAAA AAGCTCAGCC TAAACATTT ATTTCAAGCA GAAAAGAAA AGGGGGGGCG	240
60	CAAAAATGGC TGGGGCAATT ATAGAAAACA TGAGCACCAA GAAGCTGTGC ATTGTTGGTG	300

	GGATTCTGCT CGTGTTCCTAA ATCATCGCCT TTCTGGTGGG AGGCTTGATT GCTCCAGGGC	360
5	CCACAACGGC AGTGTCTTAC ATGTGGTGA AATGTGTGGA TGCCCGTAAG AACCATCACA	420
	AGACAAAATG GTTCGTGCCT TGGGGACCCA ATCATTGTGA CAAGATCOGA GACATTGAAG	480
	AGGCAATTCC AAGGGAAATG GAAGCCAATG ACATCGTGTT TTCTGTTCAC ATTCCCCTCC	540
10	CCCACATGGA GATGAGTCCT TGGTTCCAAT TCATGCTGTT TATCCTGCAG CTGGACATTG	600
	CCTTCAAGCT AAACAACCAA ATCAGAGAAA ATGCAGAACT CTCCATGGAC GTTCCCTGG	660
15	CTTACCGTGA TGACGCATTT GCTGAGTGGG CTGAAATGGC CCATGAAAGA GTACCACGGA	720
	AACTCAAATG CACCTTCACA TCTCCCAAGA CTCCAGAGCA TGAGGGCCGT TACTATGAAT	780
	GTGATGTCTT TCCTTTCATG GAAATTGGGT CTGTGGCCCA TAAGTTTAC CTTTAAACA	840
20	TCCGGCTGCC TGTGAATGAG AAGAAGAAAA TCAATGTGGG AATTGGCGAG ATAAAGGATA	900
	TCCGGTTGGT GGGGATCCAC CAAAATGGAG GCTTCACCAA GGTGTGGTTT GCCATGAAGA	960
25	CCTTCCTTAC GCCCAGCATC TTCATCAATTA TGGTGTGGTA TTGGAGGAGG ATCACCATGA	1020
	TGTCCCGACC CCCAGTGCTT CTGGAAAAAG TCATCTTTCG CCTTGGGATT TCCATGACCT	1080
	TTATCAATAT CCCAGTGGAA TGGTTTTCCA TCGGGTTTGA CTGGACCTGG ATGCTGCTGT	1140
30	TTGGTGACAT CCGACAGGGC ATCTTCTATG CGATGCTTCT GTCCTTCTGG ATCATCTTCT	1200
	GTGCGGAGCA CATGATGGAT CAGCAGGAGC GGAACCACAT TGCAGGTAT TGAAGCAAG	1260
35	TGGACCCAT TGCCGTGGC TCCTTCTGCC TCTTCATATT TGACATGTGT GAGAGAGGGG	1320
	TACAATCAC GAATCCCTTC TACAGTATCT GGAATACAGA CATTGGAACA GAGCTGGCCA	1380
	TGGCCTTCAT CATCOTGGCT GGAATCTGCC TCTGCTCTA CTTCTGTCTT CTATGCTTCA	1440
40	TGGTATTICA GGTGTTTCGG AACATCAGTG GGAAGCAGTC CAGCCTGCCA GCTATGAGCA	1500
	AAGTCCGGCG GCTACACTAT GAGGGGCTAA TTTTATAGTT CAAGTTCTTC ATGCTTATCA	1560
45	CCTTGGCCTG CGCTGCCATG ACTGTCATCT TCTTCATCGT TAGTCAGGTA ACGGAAGGCC	1620
	ATTGGAAATG GGGCGGCGTC ACAGTCCAAG TGAACAGTGC CTTTTTCACA GGCATCTATG	1680
	GGATGTGGAA TCTGTATGTC TTTGCTCTGA TGTTCTTGTA TGCACCATCC CATAAAAACT	1740
50	ATGGAGAAGA CCAGTCCAAT GGAATGCAAC TCCCATGTAA ATCGAGGGAA GATTGTGCTT	1800
	TGTTTGTTC GGAACCTTAT CAAGAATTGT TCAGCGCTTC GAAATATTC TTCAATCAATG	1860
55	ACAACGCAGC TTCTGGTATT TGAGTCAACA AGGCAACACA TGTATTATCAG CTTTGCAATT	1920
	GCAGTTGTCA CAGTCACATT GATTGTACTT GTATACGCAC ACAAATACAC TCATTTAGCC	1980
	TTTATCTCAA AATGTTAAAT ATAAGGAAAA AAGCGTCAAC AATAAATATT CTTGAGTATA	2040
60	AAAAAAAAA AAAAAAAAAA AAAAAA	2066

## 5 (2) INFORMATION FOR SEQ ID NO: 106:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1705 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

15 AATTCGGCAK AGGGCAGCTG TCGGCTGGAA GGAAGTGGTC TGCTCAGACT TGCTGGCTTG 60  
 CGCATCAGGA CTGGCTTTAT CTCTGACTC ACGGTGCAAA GGTGCACTCT GCGAACGTTA 120  
 AGTCCGTCCC CAGCGCTTGG AATCCTACGG CCCCCACAGC CGGATCCCTT CAGCCTTCCA 180  
 20 GGTCTCTAAC TCCCGTGGAC GCTGAACAAT GGCCTCCATG GGGCTACAGG TAATGGGCAT 240  
 CCGGCTGGCC GTCTGGGGCT GGCTGGCCGT CATGCTGTGC TGGCGCTGC CCATGTGGCG 300  
 25 CGTGACGGCC TTCATCGSCA GCAACATTGT CACCTCGCAG ACCATCTGGG AGGGCCTATG 360  
 GATGAAGTGC GTGGTGCAGA GCACCGGCCA GATGCACTGC AAGGTGTAGG ACTCGCTGCT 420  
 GGCAGTGGCG CAGGACCTGC AGCGGGCCCG CGCCCTCGTC ATCATCAGCA TCATCGTGGC 480  
 30 TGCTCTGGGC GTGCTGCTGT CCGTGGTGGG GGGCAAGTGT ACCAACTGCC TGGAGGATGA 540  
 AAGCGCCAAG GCCAAGACCA TGATCGTGGC GGGCGTGGTG TTCTGTTCG CCGCCCTTAT 600  
 35 GGTGATAGTG CCGGTGTCTT GGACGGCCCA CAACATCATC CAAGACTTCT ACAATCCGCT 660  
 GGTGGCTTCC GGGCAGAAGC GGGAGATGGG TGCTCGCTC TACGTGGGT GGGCGCCTC 720  
 CGGCTCTCTG CTCTCTGGCG GGGGGCTGCT TTGCTGCAAC TGTCACCCC GCACAGACAA 780  
 40 GCCTTACTCC GCCAAGTATT CTGCTGCCCG CTCTGCTGCT GCCAGCAACT ACGTGTAAAG 840  
 TGCCACGGCT CCACTCTGTT CCTCTCTGCT TTGTTCTTCC CTGGAAGTGG CTCAGCGCAG 900  
 45 GCTGTGACCC CAGGAGGGCC CTGCCACGGG CCACTGGCTG CTGGGGACTG GGGACTGGGC 960  
 AGAGACTGAG CCAGGCAGGA AGGCAGCAGC CTTACAGCTC TCTGGCCAC TCGGACAACT 1020  
 TCCCAAGGCC GCCTCTGCT AGCAAGAACA GAGTCCACCC TCCTCTGGAT ATGGGGAGG 1080  
 50 GACGGAAGTG ACAGGGTGTG GTGGTGGAGT GGGGAGCTGG CTTCTGCTGG CCAGGATGGC 1140  
 TTAACCTTGA CTTTGGGATC TGCTGCAATC GGTGTTGGCC ACTGTCCCCA TTACATTITT 1200  
 55 CCCCAGTCTG TCTGCTGCA TCTCTCTGT TCGGGTAGG CCTTGATATC ACCTCTGGGA 1260  
 CTGTGCCTTG CTCACCGAAA CCGCGGCCCA GGAGTATGGC TGAGGCCCTG CCGACCCACC 1320  
 60 TGCTTGGGAA GTGCAGAGTG GATGGACGGG TTTAGAGGGG AGGGGCGAAG GTGCTGTAAA 1380

- 358

CAGGTTTGGG CAGTGGTGGG GGAGGGGGCC AGAGAGGGCG CTCAGGTTGC CCAGCTCTGT 1440  
 GGCCTCAGGA CTCTCTGCCT CACCCGCTTC AGCCCGAGGC CCCTGGAGAC TGATCCCCCTC 1500  
 5 TGAGTCCCTCT GCCCCCTCCA AGGACACTAA TGAGCCTGGG AGGGTGGCAG GGAGGAGGGG 1560  
 ACAGCTTCAC CCTTGAAGT CTTGGGGTTT TTCTCTTCC TTCTTTGTGG TTCTGTMTT 1620  
 10 GTAATTAAAG AAGAGCTATT CATCACTGTA ATTATTATTA TTTTCTACAA TAAATGGGAC 1680  
 CTGTGCACAG GRAAAAAAAA AAAAG 1705

15

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1167 base pairs  
 20 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

TGCAGGAATT CGGCAGAGGT TTTCCGCTAG ACTCTGGCAG TTGGTGAGCA TCATGGCAAC 60  
 CGTTACAGCC ACAACCAAAG TCCCGGAGAT CCGTGATGTA ACAAGGATTG AGCGAATCGG 120  
 30 TGCCCACTCC CACATCCGGG GACTGGGGCT GGACGATGCC TTGGAGCCTC GGCAGGCTTC 180  
 GCAAGGCATG GTGGGTGAGC TGGCGGCACG GCGGGCGGCT GGCGTGGTCC TGGAGATGAT 240  
 CCGGGAAGGG AAGATTGCCG GTCGGGCAGT CCTTATTGCT GGCCAGCCCG GCACGGGGAA 300  
 35 GACGGCCATC GCCATGGGCA TGGCGCAGGC CTTGGGCCCT GACACGCCAT TCACAGCCAT 360  
 CGCCGGCAGT GAAATCTTCT CCCTGGAGAT GAGCAAGACC GAGGCGCTGA CGCAGGCCTT 420  
 40 CCGGCGGTCC ATCGGCGTTC GCATCAAGCA GGAGACGGAG ATCATCGAAG GGGAGGTGGT 480  
 GGAGATCCAG ATTGATCGAC CAGCAACAGG GACGGGCTCC AAGGTGGGCA AACTGACCCT 540  
 CAAGACCACA GAGATGGAGA CCATCTACGA CCTGGGCACC AAGATGATTG AKTCCCTGAC 600  
 45 CAAGGACAAG GTCCAGGCGG GGGACGTGAT CACCATCGAC AAGGCGACGG GCAAGATCTC 660  
 CAAGCTGGGC CGCTCCTTCA CACGCGCCCG CGAACTACGA CGCTATGGGC TCCCAGACCA 720  
 50 AGTTCGTGCA GTGCCCAGAT GGGGAGCTCC AGAAACGCAA GGAGGTGGTG CACACCGTGT 780  
 CCCTGCACGA GATCGACGTC ATCAACTCTC GCACCCAGGG CTTCCTGGCG CTCTTCTCAG 840  
 GTGACACAGG GGAGATCAAG TCAGAAGTCC GTGAGCAGAT CAATGCCAAG GTGGCTGAGT 900  
 55 GCGCGGAGGA GGGCAAGCGG GAGATCATCC CTGGAGTGCT GTTCATCGAC GAGGTCCACA 960  
 TGCTGGACAT CGAGAGCTTC TCCTTCTTCA ACCGGGCCCT GGAGAGTGAC ATGGCGCCTG 1020  
 60 TCCAGCAGGT CTATGGGGAT GCGGTGAGGG CTCTGGTAGC TGGTGCCCCG GATTGCGGTG 1080

ATGCCACGGT TGGTGGCCTC GTGCCGAATT CCTGCAGCCC GGGGGATCCA CTAGTTCTAG 1140  
AGCGGGCCGCC ACCGCGGTGG ANCTCCN 1157

5

(2) INFORMATION FOR SEQ ID NO: 108:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1907 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

20

GGCACAGGGG AATCATCGTG TGATGTGTGT GCTGCCCTTTG TGAGTGTGTG GAGTCCTGCT 60

CAGGTGTTAG GTACAGTGTG TTTGATCGTG GTGGCTTGAG GGAACCCCTT GTTCAGAGCT 120

GTGACTGCGG CTGCACTCAG AGAAGCTGCC CTTGCTGCT CGTAGCGCGG GGCCTTCTCT 180

25

CCTCGTCATC ATCCAGAGCA GCCAGTGTCC GGGAGGCAGA AGGTACCGGG GCAGCTACTG 240

GAGGACTGTG CCGGCCTGCC TGGGCTGCCC CCTCCGCCGT GGGGCCCTGT TGCTGCTGTC 300

30

CATCTATTTT TACTACTCCC TCCCAAATGC GGTGCGCCCG CCCTTCACTT GGATGCTTGC 360

CCTCCTGGGC CTCTCGCAGG CACTGAACAT CCTCCTGGGC CTCAGGGGCC TGGCCCCAGC 420

TGAGATCTCT GCACTGTGTG AAAAAGGGAA TTTCACGTG GCCCATGGGC TGGCATGGTC 480

35

ATATTACATC GGATATCTGC GGCTGATCCT GCCAGAGCTC CAGGCCCGGA TTGGAACCTA 540

CAATCAGCAT TACAACAACC TGCTACGGGG TGCAGTGAGC CAGCGGCTGT ATATTCTCCT 600

40

CCCATTGGAC TGTGGGGTGC CTGATAACCT GAGTATGGCT GACCCCPACA TTGCTTCTCT 660

GGATAAACTG CCCCAGCAGA CCGGTGACCG TGCTGGCATC AAGGATCGGG TTTACAGCAA 720

CAGCATCTAT GAGCTTCTGG AGAACGGGCA GCGGGCGGGC ACCTGTGTCC TGGAGTACGC 780

45

CACCCCTTGT CAGACTTTGT TTGCCATGTC ACAATACAGT CAAGCTGGCT TTAGCGGGGA 840

GGATAGCCTT GAGCAGGCCA AACTCTTCTG CCGGACACTT GAGGACATCC TGGCAGATGC 900

50

CCCTGAGTCT CAGAACAAC TGGCCTCAT TGCTACCAG GAACCTGCAG ATGACAGCAG 960

CTTCTCGCTG TCCCAGGAGG TTCTCCGGCA COTGCGGCAG GAGGAAAAGC AAGAGCTTAC 1020

TGTGGGCAGC TTGAAGACCT CAGCGGTGCC CAGTACCTCC ACGATGTCCC AAGAGCCTGA 1080

55

GCTCCTCATC AGTGAATGC AAAAGCCCTT CCTCTCCGC ACGGATTTCT CTTGAGACCC 1140

AGGGTCACCA GGCCAGAGCC TCCAGTGGTC TCCAAGCCTC TGGACTGGGG GCTCTCTTCA 1200

60

GTGGCTGAAT GTCCAGCAGA GGTATTTCTT TCCACACGGG GCCTTGCAGG GAAGGGTCCA 1260

GGACTTGACA TCTTAAGATG CGTCTTGTCC CCTTGGGCCA GTCATTTCCTC CTCTCTGAGC 1320  
 CTCGGTGTCT TCAACCTGTG AAATGGGATC ATAATCACTG CCTTACCTCC CTCACGGTTG 1380  
 5 TTGTGAGGAC TGAGTGTGTG GAAGTTTTTC ATAAACTTTG GATGCTAGTG TACTTAGGGG 1440  
 GTGTGCCAGG TGTCTTTTCAT GGGGCCTTCC AGACCCACTC CCCACCCTTC TCCCCTTCCT 1500  
 10 TTGCCCCGGG ACGCCGAAC TCTCAATGG TATCAACAGG CTCCTTCGCC CTCTGGCTCC 1560  
 TGGTCATGTT CCATTATTGG GGAGCCCCAG CAGAAGAATG GAGAGGAGGA GGAGGCTGAG 1620  
 TTTGGGGTAT TGAATCCCCC GGCTCCACCC CTGCAGCATC AAGGTGCTA TGGACTCTCC 1680  
 15 TGCCGGGCAA CTCTTGCCTA ATCATGACTA TCTCTAGGAT TCTGGCACCA CTTCTTCCC 1740  
 TGGCCCCCTTA AGCCTAGCTG TGTATCGGCA CCCCCACCCC ACTAGAGTAC TCCCTCTCAC 1800  
 TTGCGGTTTC CTTATACTCC ACCCCTTTCT CAACGGTCCT TTTTAAAGC ACATCTCAGA 1860  
 20 TTAACAAAAA AAAAAA AAAAAGGG CGGCCGC 1907

25

(2) INFORMATION FOR SEQ ID NO: 109;

(i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 611 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

35

40

45

50

55

60

ATGAATTAAC GCCAAGCTNT NAATAGGGAC TCACTATGGG GGAAAGNTGG GTAACGCCTG 60  
 CAGGTACCGT TCCGGAATTC CCGGGTCGAC CCACGCGTCC GATGGGGCTT TAGTAAATCA 120  
 GGCTTGCAGG CTCAAAGCTG CAATCTGCCC ACTCTCAGGT ACTGAGACTT TGTGGGCCTC 180  
 AGACACCAGG AAGAAAGTTG GGATACAGTC ATTTGAGTTA AAAAGGGAAT GACCCCTCAG 240  
 AAACCCGCAT TAGCAGTGT ACTCTTGAA GTGCCTTTAC TTTAACGCT CTCTGTTCTG 300  
 AAAAAGAGGT GTTTGGTTAC GTGTAGCCA ACATCACGTT TTGTTAGCTG TGATTACCT 360  
 TTGTCCGTTT AAAAGACTTC ACGGAGCCAT TCTGTATACA AGGTGTGCTC TTTCCAATGT 420  
 AGAAGGGGTT ATGGAAAAGG GTCCGATCCT TTGCTGTAAA CTGGAGAGAC CAGTCCCAAA 480  
 CAGAGGGGAA TTTAAGCCC TTCTCATCAC CCAATTGGAT GTTTTTCCTT ATAGCAAATT 540  
 CCTGCAAAAT AAATAAATAA ATATTTGCAA AACTAAAAA AAAAAAAAAA AAAAAAAAAA 600  
 GGGGGGNCN C 611



## (2) INFORMATION FOR SEQ ID NO: 110:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2632 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

10 TCCCAGCTCT CAGGACAAGG GCCCTGGGCG ATCTTTTAAA AAAGCCGATT GGGTGTCTTT 60  
 CTAAAANTAC AACCACTACT TCATCGTCAA GTTCTGCGGA AGGAGTCCC CTCCAGATTC 120  
 15 TCATGGAGTG ACAAATCTTG ACTCTTGCTC CTGGAATTTT TCAGGCCCAA ACTAGCGTTT 180  
 CTACAATGAT TTATTTGGCA AATTGTCTT GATTATGGGT GGCTGATGAG GAACGTGCTT 240  
 20 TTGTTAGGAA CCGAAACTGG GCGCGGTGA GGGCGTGAC GCAATGAGTC CGGAAGAGGG 300  
 TGAAATGCTT TCGGTAGGCA CTCACGGCT GTGAAGATGG CCGCGCTGC GTGGCTTCAG 360  
 GTGTTGCGTG TCATTCTCT GCTCTGCGA GCTCACCCT CACCACTGTC GTTTTTCAGT 420  
 25 GCGGACCGG CAACCGTAGC TGCTCCGAC CGGTCCAAAT GCCACATTCC GATACCGTCG 480  
 GGGAAAAATT ATTTAGTTT TGGAAAGATC CTCTTCAGAA ATACCACTAT CTTCTGAAG 540  
 TTTGATGGAG AACCTTGTA CTTGTCTTTG AATATAACCT GGTATCTGAA AAGCGCTGAT 600  
 30 TGTTACAATG AATCTATAA CTCAAGGCA GAAGAAGTAG AGTTGTATTT GGAATAACTT 660  
 AAGGAAAAAA GAGGCTTGTC TGGAAATAT CAAACATCAT CAAAATTGTT CCAGAACTGC 720  
 35 AGTGAACCTT TAAAACACA GACCTTTTCT GGAGATTTTA TGCATCGACT GCCTCTTTTA 780  
 GGAGAAAAAC AGGAGCTAA GGAGAATGGA ACAAACCTTA CCTTTATTGG AGACAAAACC 840  
 GCAATGCATG AACCATGCA AACTTGGCAA GATGCACCAT ACATTTTAT TGTACATATT 900  
 40 GGCATTTTAT CCTCAAAGG ATCATCAAAA GAAAATTCAC TGAGTAATCT TTTTACCATG 960  
 ACTGTTGAAG TGAAGGTCC CTATGAATAC CTCACACTTG AAGACTATCC CTTGATGATT 1020  
 45 TTTTTCATGG TGATGTGAT TGTATATGTC CTGTTGGTG TTCTGTGGCT GGCATCGTCT 1080  
 GCCTGCTACT GGAGAGATCT CTTGAGAATT CAGTTTGGG TTGGTGCTGT CATCTTCCTG 1140  
 GGAATGCTTG AGAAAGCTGT CTTCTATGCG GAATTCAGA ATATCCGATA CAAAGGARA 1200  
 50 TCTGTCCAGG GTGCTTTGAT CTTGTCAGG CTGCTTTCAG CAGTGAAACG CTCACTGGCT 1260  
 CGAACCCCTG TCATCATAGT CAGTCTGGG TATGGCATCG TCAAGCCACG CCTGGACTCA 1320  
 55 CTCTTCATAA GGTGTAGTA GCAGRAGCCC TCTATCTTTT GTTCTCTGGC ATGGAAGGGG 1380  
 TCCTCAGAGT TACTGGGGCC CAGACTGATC TTGCTTCCTT GGCCTTTATC CCCTTGGCTT 1440  
 60 TCCTAGACAC TGCCTTGTC TGGTGGATAT TTATTAGCCT GACTCAAACA ATGAAGCTAT 1500

362

TAAAACTTCG GAGGAACATT GTAAAACTCT CTTTGTATCG GCATTTCCACC AACACGCTTA 1560  
 TTTTGGCAGT GGCAGCATCC ATTGTGTTTA TCATCTGGAC AACCATGAAG TTCAGAATAG 1620  
 5 TGACATGTCA GTCGGACTCG CGGGAGCTGT GGGTAGACGA TCCCATCTGG CGCTTGCTGT 1680  
 TCTCCATGAT CCTCTTTGTC ATCATGGTTC TCTGGCGACC ATCTGCAAAC AACCAAGAGT 1740  
 10 TTGCCTTTTC ACCATTGTCT GAGGAAGAGG AGGAGGATGA ACAAAGGAG CCTATGCTGA 1800  
 AAGAAAGCTT TGAAGGAATG AAAATGAGAA GTACCAAACA AGAACCCAAT GGAAATAGTA 1860  
 AAGTTAACAA AGCACAGGAA GATGATTTGA AGTGGGTAGA AGAGAATGTT CCTTCTTCTG 1920  
 15 TGACAGATGT AGCACTTCCA GCCCTTCTGG ATTGAGATGA GGAACGAATG ATCACACACT 1980  
 TTGAAGGTC CAAAATGGAG TAAGGAATGG GAAGATTGTC AGTTAAAGAT GGCTACCATC 2040  
 AGGGAAGAGA TCAGCATCTG TGTCACTCTT CTGTACGGCT CCATGGGATT AAAGGAAGCA 2100  
 20 ATGACATCCT GATCTGTTCC TTGATCTTTG GGCATTGGAG TTGGCGAGAG GTGTCAAGAC 2160  
 AAAGAGAACA TCTTACTGAA AACAAGTTCA TAAGATGAGA AAAATCTACG AGCTTCTTAT 2220  
 25 TTACAACACT GCTGCCCCCT TTCCTCCAG ACTCTGACAT GGATGTTTAT GCAACTTAAG 2280  
 TGTGTGTTTC CTGAACTTTC TGTAATGTTT CATTTTTAA ATCTGACAAA CTAAAAAGTT 2340  
 TAACGTCTTC TAAAAGATTG TCATCAACAC CATAATATGT AATCTCCAGG AGCAACTGCC 2400  
 30 TGTAATTTTT ATTTATTTAG GGAGTTACAT AGGTGATGGG GGAAATTGTT AACTACCTTT 2460  
 CATTTCCTG GGAAGTCAAG GTTACATCTT GCAGAGGTTG TTTTGAGAAA AAAGGGCCCT 2520  
 35 TCTGAGTTAA GGAGCCATAG TTCTATCAAT GATCAAAAGA AAAAAAAAAA AACTCGATCG 2580  
 GCACGAGGGG GGGCCCCGTA CCCAATTCGC CCTATGGGAN TCGAATGAGA CC 2632

40

(2) INFORMATION FOR SEQ ID NO: 111:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2249 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

GAATTCGGCA CGAGCTCACC GTGCTGCGTG ACACAAGGCC AGCCTGCGCC TACGAGCCCA 60  
 TGGACTTTKT RATGGCCCTC ATCTACGACA TGGTACTGSW TGTGGTCACC CTGGGGCTGG 120  
 55 CCCTCTTCAC TCTGTGCGGC AAGTTCAAGA GGTGGAAGCT GAACGGGGCC TTCCTCCTCA 180  
 TCACAGCCTT CCTCTCTGTG CTCATCTGGG TGGCCTGGAT GACCATGTAC CTCTTGGCA 240  
 60 ATGTCAAGCT GCAGCAGGGG GATGCCTGGA ACGACCCAC CTTGGCCATC ACGCTGGCGG 300

	CCAGCGCTGG GTCTTCGTCA TCTTCCACGC CATCCCTGAG ATCCACTGCA CCTTCTGCG	360
5	AGCCCTGCGG GAGAACACGC CCACTACTT CGACAGCTCG CAGCCCAGGA TCGGGGAGAC	420
	GGCCTTCGAG GAGGACGTGC AGCTGCCGCG GGCCTATATG GAGAACAAGG CCTTCTCCAT	480
	GGATGAACAC AATGCAGCTC TCCGAACAGC AGGATTTCCT AACGGCAGCT TGGGAAAAAG	540
10	ACCCAGTGGC AGCTTGGGGA AAAGACCCAG CGCTCCCTTT AGAAGCAACG TGTATCAGCC	600
	AACTGAGATG CCCGTCGTGC TCAACGGTGG GACCATCCCA ACTGCTCCGC CAAGTCACAC	660
15	AGGAAGAMAC CTTTGGTGAA AGACTTTAAG TTCCAGAGAA TCAGAAATTC TCTTACCGAT	720
	TTGCCTCCCT GGCTGTGTCT TTCTTGAGGG AGAAATCGGT AACAGTTGCC GAACCAGGCC	780
	GCCTCACAGC CAGGAAATTT GGAAATCCTA GCCAAGGGGA TTTCGTGTAA ATGTGAACAC	840
20	TGACGAACTG AAAAGCTAAC ACCGACTGCC CGCCCTCCCT CTGCCACACA CACAGACAGC	900
	TAATACCAGA CCAACCTCAA TCCCGCAAAA CTAAAGCAAA GCTAATTGCA AATAGTAATA	960
25	GGCTCACTGG AAAATGTGGC TGGGAAGACT GTTTCATCCT CTGGGGGTAG AACAGAACCA	1020
	AATTCACAGC TGGTGGGCCA GACTGGTGTT GGTGGAGGT GGGGGGCTCC CACTCTTATC	1080
	ACCTCTCCCT AGCAAGTGCT GGACCCGAGG TAGCCTCTTG GAGATGACCG TTGCGTTGAG	1140
30	GACAAATGGG GACTTTGCCA CCGGCTTTGC CTGGTGGTTT GCACATTTCA GGGGGGTGAG	1200
	GAGAGTTAAG GAGGTGTGG GTGGGATTCC AAGGTGAGGC CCAACTGAAT CGTGGGGTGA	1260
35	GCTTTATAGC CAGTAGAGGT GGAGGGACCC TGGCATGTGC CAAAGAAGAG GCCCTCTGGG	1320
	TGATGAAGTG ACCATCACAT TTGGAAAGTG ATCAACCACT GTTCCTTCTA TGGGGCTCTT	1380
	GCTCTAGTGT CTATGGTGAG AACACAGGCC CGGCCCTTC CCTTGTAGAG CCATAGAAAT	1440
40	ATTCTGGCTT GGGGCAGCAG TCCCTTCTTC CCTTGATCAT CTGGCCCTGT TCCTACACTT	1500
	ACGGGTGTAT CTCCAAATCC TCTCCCAATT TTATTCCCTT ATTCATTTCA AGAGCTCCAA	1560
45	TGGGGTCTCC AGCTGAAANS CCCTCCGGGA GGCAGGTTGG AAGGCAGGCA CCACGGCAGG	1620
	TTTTCCGCGA TGAATGCACC TAGCAGGGCT TCAGGGGTTT CCACTAGGAT GCAGAGATGA	1680
	CCTCTCGCTG CCTCACAAGC AGTGACACCT CGGGTCCTTT CCGTTGCTAT GGTGAAAATT	1740
50	CCTGGATGGA ATGGATCACA TGAGGTTTTC TTGTTGCTTT TGGAGGGTGT GGGGGATATT	1800
	TTGTTTTGGT TTTTCTGCG GTTCCATGAA AACAGCCCTT TTCCAAGCCC ATTGTTTCTG	1860
55	TCATGGTTTC CATCTGTCTT GAGCAAGTCA TTCCTTTGTT ATTTAGCATT TCGAACATCT	1920
	CGGCCATTCA AAGCCCCCAT GTTCTCTGCA CTGTTTGGCC AGCATAACCT CTAGCATCGA	1980
	TTCAAAGCAG AGTTTAAACC TGACGGCATG GAATGTATAA ATGAGGGTGG GTCCTTCTGC	2040
60	AGATACTCTA ATCACTACAT TGCTTTTTCT ATAAACTAC CCATAGCCT TTAACCTTTA	2100

AAGAAAAATG AAAAAGTTA GTTTTGGGG GCGGGGGGAG GACTGACCGC TTCATAAGCC 2160  
 AGTACGTCTG AGCTGAGTAT GTTCAATAA ACTTTTGAT ATTTCTCAA AAAAAAAAAA 2220  
 AAAAAACCCG GCGGGGGGCG GCGACTGG 2249

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2193 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

GATACTATTA GCGAGTGC TCGGGTGC GCGGTAGAC TAGTGGATCC CGGGTGCAGG 60  
 AATTCGGCG AGCGCGCGG GAGCGGAGT GGTGGCGCGC CGCGCGCGCG TGCGTCCGGG 120  
 GANCCAAAA TCGTGAAT CAGCTGAGG ACCCGGAACA AAAGGAGGAA TTCGCCGTGC 180  
 CCGAGATAG CTCCGTCCG CAGTTAAGG AAGAAATCTC TAAACGTTTT AAATCACATA 240  
 CTGACCAACT TGTGTGATA TTTGTGGA AAATTTTGAA AGATCAAGAT ACCTTGAGTC 300  
 AGCATGGAAT TCATGATGA GTTACTGTC ACCTTGTCAT TAAAACACAA AACAGGCCTC 360  
 AGGATCATTC AGCTCAGTA ACAATACAG CTGGAAGCAA TGTTACTACA TCATCAACTC 420  
 CTAATAGTAA CTCTACATT GTTCTGCTA CTAGCAACCC TTTTGGTTTA GGTGGCCTTG 480  
 GGGGACTTGC AGGTGTGAT AGCTTGGTT TGAATACTAC CAACTTCTCT GAACTACAGA 540  
 GTCAGATGCA GCGCAACTT TTTTAAACC CTGAATGAT GGTCCAGATC ATGGAAAAC 600  
 CCYTTGTTCA GAGCATGTC ATCAATGCT GACCTGATG AGACAGTTAA TTATGGCCAA 660  
 TCCACAAATG CAGCGTTTA TACGAGAAA TCCAGAAAT TAGTCATATG TTGAATAATC 720  
 CAGATATAAT GAGACAAAG TTGAACTTG CCCAGGAATC CAGCAATGAT GCAGGAGATG 780  
 ATGAGGAACC AGGACCGAG TTTGAGCAAC CTAGAAAGCA TCCCAGGGGG ATATAATGCT 840  
 TTAAGGCGCA TGTACACGA TATTCAGGA CCAATGCTGA GTGCTGCACA AGACGAGTTT 900  
 GGTGGTAATC CATTTGCTTC CTGCTGAGC AATACATCCT CTGGTGAAGG TAGTCAACCT 960  
 TCCGTACAG AAAATAGAA TCCACTACCC AATCCATGGG CTCCACAGAC TTCCAGAGT 1020  
 TCATCAGCTT CAGCGGGAC TGGCGCACT GTGGGTGGCA CTAAGGTAG TACTGCCAGT 1080  
 GGCATTTCTG GGCAGATAC TACTGCGCA AATTTGCTGC CTGGAGTAGG AGCTAGTATG 1140  
 TTCACACAC CAGGATGCA GAGTTGTTG CAACAATAA CTGAAACCC ACAACTTATG 1200

365

CAAAACATGT TGTCTGCCCC CTACATGAGA AGCATGATGC AGTCACTAAG CCAGAATCCT 1260  
 GACCTTGCTG CACAGATGAT GCTGAATAAT CCCCTATTTG CTGGAAATCC TCAGCTTCAA 1320  
 5 GAACAAATGA GACAACAGCT CCCAACTTTC CTCCACAAA TGCAGAATCC TGATACACTA 1380  
 TCAGCAATGT CAAACCCTAG AGCAATGCAG GCCTTGTTAC AGATTCAGCA GGGTTTACAG 1440  
 10 ACATTAGCAA CGGAAGCCCC GGGCCTCATC CCAGGGTTTA CTCCTGGCTT GGGGGCATT 1500  
 GGAAGCACTG GAGGCTCTTC GGGAACTAAT GGATCTAAG CCACACCTAG TGAAAACACA 1560  
 AGTCCCACAG CAGGAACCAC TGAACCTGGA CATCAGCAGT TTATTCAGCA GATGCTGCAG 1620  
 15 GCTCTTGCTG GAGTAAATCC TCAGCTACAG AATCCAGAAG TCAGATTTCA GCAACAACCTG 1680  
 GAACRACTCA GTGCAATGGG ATTTTTGAAC CGTGAAGCAA ACTTGCAAGC TCTAATAGCA 1740  
 ACAGGAGGTG ATATCAATGC AGCTATTGAA AGGTTACTGG GCTCCAGCC ATCATAGCAG 1800  
 20 CATTTCTGTA TCTKGA AAAA ATGTAATTTA TTTTGTATAA CGGCTCTTAA ACTTTAAAAT 1860  
 ACCTGCTTTA TTTCAATTTG ACTCTTGGA TTTGTGCTG TTATAAACAA ACCCAATATG 1920  
 25 ATGCATTTTA AGGTGGAGTA CAGTAAGATG TGTGGGTTTT TCTGTATTTT TCTTTTCTGG 1980  
 AACAGTGGGA ATTAAGGCTA CTGCATGCAT CACTTCTGCA TTTATTGTAA TTTTTTAAAA 2040  
 ACATCACCTT TTATAGTTGG GTGACCAGAT TTTGTCTGCT ATCTGTCCAG TTTATTTGCT 2100  
 30 TTTTAAACAT TAGCCTATGG TACTAATTTA TGTAGAATAA AAGCATTAAA AAGAAGCAA 2160  
 AAAAAAAAAA AAAAATTCCT GCGCCCGCGA ATTCTTCT 2198

## (2) INFORMATION FOR SEQ ID NO: 113:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1043 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

CTGAAGTGTA TGTGGTGAGG AAGAAGAGGC TCCTACTGTA GACAGCCTTG TTCTACAGAT 60  
 50 CCTCCCAGAA ATCTCTGGGC CAGGTGGAAC CCAGGGTCAG AGAGGGATGG GAGAGAGGTT 120  
 TAATTTTCCA TGATAAATAA AAATCTATAA AATAATAAAC AAGAGAAAAG AGATTGGAAA 180  
 55 CAGCCAGGTT GGAGCAGTGA GTGAGTAAGG AAACCTGGCT GCGCTCTCCA GATTCCCCAG 240  
 GCTCTCAGAG AAGATCAGCA GAAAGTCTGC AAGACCCTAA GAACCATCAG CCCTCAGCTG 300  
 CACCTCTCTC CCTCCAAGGA TGACAAAGGC GCTACTCATC TATTTGGTCA GCAGCTTTCT 360  
 60 TGCCCTAAAT CAGGCCAGCC TCATCAGTCG CTGTGACTTG GCCCAGGTGC TGCAGCTGGA 420

5  
 10  
 15  
 20  
 25

RGA CT TGGAT GGG TTT GAGG GTT ACT C C C T GAG T GACT GG CTGTGCCTGG CTTTGTGGA 480  
 AAGCAAGTTC AACATATCAA AGATWAATGA AAATGCAGAT GGAAGCTTTG ACTATGGSCT 540  
 CTTCCAGATC AACAGCCACT ACTGGTGCAA CRATTATAAG AGTTACTCGG AAAACCTTTG 600  
 CCACGTAGAC TGTCAAGATC TGCTGAATCC CAACCTTCTT GCAGGCATCC ACTGCGCAA 660  
 AAGGATTGTG TCCGGAGCAC GGGGGATGAA CAACTGGGTT AGAATGGAAG KTTGCACTGT 720  
 TCAGGCGGGC CACTCTTCTA CTGGCTGACA GGATGCCGCC TGAGATKAAA CARGGTCCGG 780  
 GTGCACCGTG GARTCAATCC AAGACTCCTG TCCTCACTCA RGGATTCTTC ATTTCTTCTT 840  
 CCTACTGCCT CCACTTCATG TTATTTTCTT CCGTTCCCAT TTACAACTAA AACTGACCAG 900  
 AGCCCCAGGA ATAAATGGTT TTCTTGGCTT CCTCCTTACT CCCATCTGGA CCCAGTCCCC 960  
 TGGTTCTGT CTGTTATTTG TAAACTGAGG ACCACAATAA AGAAATCTTT ATATTTATCG 1020  
 AAAAAAAAAA AAAAAAACT CGA 1043

(2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 703 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

35  
 40  
 45  
 50  
 55  
 60

GAATTCGGCA CGAGTGGCG GGCACCACGG CGGTTTTTCG ACGCTGCCGG TGGACGCAGG 60  
 CAGCATGGAC CACGGTTGCT GGGCGGATGG GGAGCGTCTA TGGTCAGTTG CCTTAGAAGT 120  
 GGTGAGATGG GAAGCTGCAG TTGGAAGACC CTGGAGGATG CCTGACAAGG GGATGTCTGA 180  
 CACATGATTG GAGCTCTTTT TGAAATGTTT CTGCCCCTTC CTGGAGCAGA GGAGCCATTA 240  
 TTTATGCAGG TACATCGAAG TCTTTTGACC TCCATACAGT GATTATGCTT GTCATCGCTG 300  
 GTGGTATCCT GCGGGCCTTG CTCCTGCTGA TAGTTGTGCT GCTCTGTCTT TACTTCAAAA 360  
 TACACAACGC GCTAAAAGCT GCAAAGGAAC CTGAAGCTGT GGCTGTAAAA AATCACAACC 420  
 CAGACAAGGT GTGGTGGGCC AAGAACAGCC AGGCCAAAAC CATTGCCACG GAGTCTTGTC 480  
 CTGCCCTGCA GTGCTGTGAA GGATATAGAA TGTGTGCCAG TTTGATTCC CTGCCACCTT 540  
 GCTGTTGCGA CATAAATGAG GGCCTCTGAG TTAGGAAAGG TGGGCACAAA AATCTTCATG 600  
 AGCAATACTT CTTAGTAGAT TGTTTTGTTA TTCAAATCAA GTTCTAGTGT TTTTATGTGA 660  
 GATTATATAA TTTACAGTGT TGTTTTATAT ACTTTTGAAT AAA 703

## (2) INFORMATION FOR SEQ ID NO: 115:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3684 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

5	GGCAGAGGGG GCATGAGCAG GAGGAGGATT ACCGCTACGA GGTGCTCAGG GCCGAGCAGA	60
15	TTCTACAACA CATGGTGGNA ATGTATCCGG GAGGTCAACG AGGTCATCCA GAATCCAGCA	120
	ACTATCACAA GAATACTCCT TAGCCACTTC AATTGGGATA AAGAGAAGCT AATGGAAGG	180
20	TACTTTGATG GAAACCTGGA GAAGCTCTTT GCTGAGTGTC ATGTAATTAA TCCAAGTAAA	240
	AAGTCTCGAA CAGCCAGAT GAATACAAGG TCATCAGCAC AGGATATGCC TTGTCAGATC	300
	TGCTACTTGA ACTACCCTAA CTCGTATTC ACTGGCCTTG AATGTGGACA TAAGTTTGT	360
25	ATGCAGTGCT GGAGTGAATA TTAACTACC AAAATAATGG AAGAAGGCAT GGGTCAGACT	420
	ATTTCTGTGC CTGCTCATGG TTGTGATATC TTAGTGGATG ACAACACAGT TATGCGCCTG	480
30	ATCACAGATT CAAAAGTTAA ATTAAAGTAT CAGCATTTAA TAACAAATAG CTTTGTAGAG	540
	TGCAATCGAC TGTTAAAGTG GTGTCTGCC CCAGATTGCC ACCATGTGT TAAAGTCCAA	600
35	TATCCTGATG CTAAACCTGT TCGCTGCAAA TGTGGGCGCC AATTTTGCTT TAACTGTGGA	660
	GAAAATTGGC ATGATCCTGT TAAATGTAAG TGGTTAAAGA AATGGATTAA AAAGTGTGAT	720
	GATGACAGTG AAACCTCCAA TTGGATTGCA GCCAACACAA AGGAATGTCC CAAATGCCAT	780
40	GTCACAATTG AGAAGGATGG TGGTTGTAAT CACATGGTCT GTCGTAACCA GAATTGTAAA	840
	GCAGAGTTTT GCTGGGTGTG TCTTGGCCCA TGGGAACCAC ATGGATCTGC CTGGTACAAC	900
45	TGTAACCGCT ATAATGAGGA TGATGCAAAG GCAGCAAGAG ATGCACAGGA GCGATCTAGG	960
	GCAGCCCTGC AGAGGTACCT GTTCTACTGT AATCGCTATA TGAACCACAT GCAGAGCCTG	1020
	CGCTTTGAGC ACAAACCTATA TGCTCAGGTG AAACAGAAAA TGGAGGAGAT GCAGCAGCAC	1080
50	AACATGTCTT GGATTGAGGT GCAGTTCCTG AAGAAGGCAG TTGATGTCTT CTGCCAGTGT	1140
	CGTGCCACAC TCATGTACAC TTATGTCTTC GCTTTCTACC TCAAAAAGAA TAACCAGTCC	1200
55	ATTATCTTTG AGAATAACCA AGCAGATCTA GAGAATGCCA CAGAGGTGCT CTCGGGCTAC	1260
	CTTGAACGAG ATATTTCCCA AGATTCTCTG CAGGATATAA AGCAGAAAGT ACAAGACAAG	1320
	TACAGATACT GTGAGAGTCG ACGAAGGGTT TTGTTACAGC ATGTGCATGA AGGCTATGAA	1380
60	AAAGATCTGT GGGAGTACAT TGAGGACTGA GAATGGCCCT GCATAAATG AACTCTGAAA	1440

	ACTTTACCAT CTAGAGTGCT CATGCAATTA AAACAAAACA AACACAAACA AGGAGGCACT	1500
5	AAGCCTATTC TGACACCACT GGTCTGTAGT ACCAGAATTG TTTTGTTRAT GGAAAGTTTA	1560
	AGTAAATTAT ATTGTAATAA AAAGGTAGAT AAACCATTGT ACRACAGTAT TCTAGGCCGC	1620
	CAACAAAAGT GTGACAGACA CACTAAAAGC CCTCCAACCT TAACTTGTA CGTAGCTTCA	1680
10	TTCTCAAAGC TGACTCCTTT TTTTCTTTT TCCTTTTCCT GAGTGTAGTA CAGTTAAAT	1740
	TTCAAACAGC TCCTTGACAC TGCTTTTCAT GTTCAAACCA GCCATTTTGT TGTACTTTGG	1800
15	TAAAGGACCT CTTCCCTTC CTCCCTACA CATACAGATA CACCCACACA CAGACTGACT	1860
	CTCTTTCTCT CATACCCCAA GGTCAAGT GAATGATGCT TAGTTCTTG TAAAGAAAAT	1920
	CTTGGCATGG GGAAAGGGGT AGGCAGCAAG AGGATTCAAC AAACGAAAAA CATAAAAAT	1980
20	TTGTATATGA CTTTAAAC AAGAGGACAA CACAGTATTT TTCAAAATTG TATATAGCGC	2040
	ATATGCATGG ACAAGCAAG CGTGGCACGT GTTGCATAA TGTTTAATTA CAAAAAATA	2100
25	TTTATTCTTT AAAAATCTTC AAGATTATGT CTATTGCTG TCCATTTTCT TTCAGTTTGC	2160
	TTATCTTTCC CGGGTGGGG TTGGGATAAA GGTGTGTCGG TTAGCACCT CTGAAGACC	2220
	TATCTAGAGC TCCTTCACTT TCCTGAGGTT ATTTTGCCCT TCTGGTGTG GGTATGCTG	2280
30	TTGCCGGCCA TGGGCTNCAY GCCTTGAATT CCTGCTCTG ATCAGGGACA AGGGAGGTCA	2340
	AGCTCTGACT AATGCCATGA CCTGATTAAG GGTACAGCA GGGAGTTTG TTGCTACAGC	2400
35	TCATGAATTA ACCTGTCCGA ACCTAATCCC CCTCCATGGC ATCATGCCTC TACCCAAGCC	2460
	TTTGTGTGCC CATGTTATGC ACACAGCTGT AGGCATTCTT AAGTCCCTG TCGCATCCAG	2520
	TGGAAGCAAT TTAATAATTC TTTTACTTTT TGGTTTCCC TTAATTGCTG CTTTTCAGAT	2580
40	TTTAGTTATG GCTCGTCTGC TCACCCCTTC TCTACATTAG GGTGTCAAAG AGAATGTTTT	2640
	GCTTTAATA TAAATAGCCA TTCATTTAGT CTCAGATTGT GAATTTAAAA TGGTGGATAC	2700
45	CGAAATTGCT TGTGTGTGTT GCTGTGGGTT TGGTTTGAAG GCAAACACCC CTAGAACATG	2760
	ATATTCCCAT CTAGTGCAAT TAAATAGAAA TCACTGAGTT TGCTGCTTTT TTATTGTGAG	2820
	CAGATAGGAG AATTAATAAT GCATTTTAGC TGTGATGTCC ATTTTATGA AATTCCTACT	2880
50	AAGAGCTATG TTAAGAGTAA AGGATGGTGG TGGTTGTATT AACTATATAC CTGTTTAGGC	2940
	CATTCTGGCT GTGGTATTTT TCAATAGGTC AGCATCTGTA AATCTGTCAG TTTTATACAG	3000
55	GAGTGCAGAG TGAAGTAGGC AACTAGATTA AGAGGTCTAA ATATGAAATA CCAGTTGAGG	3060
	CTGAGGACCT CTTGCTCTTC CTTTAAATGT CTTTGGCTA GGGAGTGTG ACCATTTGTG	3120
	AGGCAGCTTT GTCTGCTCTT AACTGTACA TCCTATTACT CCATTGGGAA GTAGGTTTAC	3180
60	TTTCCTCTGG CTTTGTGCTT AAGTTAGGCT TTGCTGAATC AACCTACTT TTCCTTTTAG	3240



AAAAGGTTGT TACAGGAGAT TTAAGTGGCA CTGTTCTTTT CCCATCAAAA ATCAGTGAAT 3300  
 GTTTCGCTGAG TATAAATGCT GCTTCCTTAA ACCACTTGTG GCTTTAGGAT CAACTTTACC 3360  
 5 TGTACCTTTT CTCCTTTTCT CCCTTGCCAC CTCAGGTGCA AATCTGAAC CAGTGTCTGC 3420  
 TTCTTCCATT TTCTCGTCTC TCTCCCTCT TCCCCATT TCCATATGAC ATTATTTTAC 3480  
 10 TTCAAATGAC AGCATCAATC TTAAGGAGAT ATACATTAAA ACTAAGGAGT TTTTAAAG 3540  
 AAAGCCTGAA TAAGTTCCTT TCCCTGGTAA CTTTGAAAAG CAGTCAGAGT TCCTATATAG 3600  
 ATATATGTGG CTCCTTTAAA ATGCTTTGTG TATGTGTGGT GTTTAAAAA AAAAAAAAAA 3660  
 15 TTCGGGGGGG GGCCCGGTNC CCAT 3684

20 (2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1965 base pairs  
 25 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

30 AAGAAAGGGT ATTAAGATTC TAGATCACAT ATGGACCCGG GAAGGTTTTT NACCCCTCTGT 60  
 TAGTGACATC GAGTCTCCCA CTAGACAAA TAGGTGGAAA AATCTCTCGA GGGCTCACAT 120  
 35 TGTTTTGTC TCTTCAGGAA AAACACCACC AGGCCATACC ACAGCCTGCC CAGTGAGGGG 180  
 GTCTTTGCCA ACAGCACCGG GATGCTGGTG GTGGCCTTTG GGCTGCTGGT GCTCTACATC 240  
 CTCTGGCTT CATCTTGGA GCGCCAGAG CCGGGATCC TGACCGACAG ACAGCCCTG 300  
 40 CTGCATGATG GGGAGTGAAG CAGCAGGAAG GGGCTCCCAA GAGCTCCTGG TGGTCAGCC 360  
 TGTGCTCCCC TCAGAAGCTC TGCTCTTCCC AGGGCTCCCC GCTGGTTTCA GCAGGCGACT 420  
 45 TTCTTCCAAT GCTGGGCCCC GACTTCTTGC CTGGGTGCTG GCTGCGCTC TCCGNCGGC 480  
 TTGCTGCTG TCTGCTTCC TTGGTGGT TTGCTGGTGC TGGGCTGCC CTCTCCGGC 540  
 GCTTGCTGCC TGTCTGCTT CTTGGTGGC TTTGCTGGGT GCTGGGCTG CTTCTCTGG 600  
 50 CTGCTTGTG CTTGTCTGCT TTTCTGGTG GCTTTGGCTT CTGCACTCCT TGGCGTCAC 660  
 TCTCAGGTCC TCCATTGACA CGAGGTCTC CTCGCTCTGG CCGCTCTTGC TGCTCCTGTC 720  
 55 TGAACAWATC AGACTGATTT CTTCTAAGA CTCCTAGGGA TGTGGTGAAG AGCTGGGACT 780  
 CAAGTGCACT CCACGGTGTG AAACATGAGG GARGTGAGGT GTCCGTCCAC TTCCCCCATA 840  
 AAGGTGTGCA TTTCACTTAG GCTGCCCCGC CACAGAGCAG GCTTCATCTG CTCTGCCATC 900  
 60

CAGCCCCATC TGGATGTGAG GTGGGGTGGG GACATCATGG GGTGATTGCA GAAAGGGGGA 960  
 GTGGCGGGCC ACGCAGCTTC TGCTGAGGAG CTGACCGCTC TGAGCTGTTT TGTTCGTAT 1020  
 5 TGCTGCTCTG TGTCTGCATG TATTGTGACC GTGCGGCTCC ACCTCTTCCA GCTGCTGCTA 1080  
 CAGCTGAGGC CTGGATCCCG GCCTTTCCCT GTGACTTACG TGTCTGTAC CCGCANGCAG 1140  
 10 CCCTACAAAT CCTGGTGACC TGCTCTCCA AGAACAGAGC CTGTCCCCAG ATGTCCCACT 1200  
 AGCGATGAGT AACAGAGGTG GCTGTGGACT TCCTCTACTT CTCCTTGCTG GATCAGGGCC 1260  
 TTCCTGCCTC CCGCTGGGCA GGTCTGCCCT TGCTCTCTTG GCAGGGCCCC AGCCCCCTCTG 1320  
 15 ACCACTCTGC AGCTCACCAT GCAGCTGATG CCAAAGTTGT GGTGTCCACT GTGCAGCAGC 1380  
 CCTGGGAGCC ACTGCCACCT TCAGAGGGGT TCCTTGCTGA GACCCACATT GCTTCACCTG 1440  
 GCCCCACCAT GGCTGCTTGC CTGGCCCAAC CTAGCGTCTT GTGCCATGCT AGAGCTTGAG 1500  
 20 CTGTTGCTCT TCTTCAGGGG ACGAAATACG GTGGAGAGCG CGAAGGGTCT TGCTCCTAAG 1560  
 TGTGCTGCT GTGGCTTTTT TGCCTTCTCC AAAGACGCAC TGCCAGGTCC CAAGCTTCAG 1620  
 25 ACTGCTGTGC TTAGTAAGCA AGTGAGAAGC CTGGGGTTTG GAGCCACCT ACTCTCTGCC 1680  
 AGCATCAGCA TCCTACTCCT GGCAACATCA GGCCAACGTC CACCCAGCC TCACATTGCC 1740  
 AGATGTTGGC AGAAGGGCTA ATATTGACCG TCTTGACTGG CTGGAGCCTT CAAAGCCACT 1800  
 30 GGGATGTCTT CCAGGCACCT CGCTCCCATG ACCAGCTCCC CGTCTCCATA GGGGTAGGCA 1860  
 TTTCACTGGT TTATGAAGCT CGAGTTCAT TAAATATGTT AAGAATCAA GCTGTCTTTG 1920  
 35 TTCAGGCTGC TATAACAAA ATATAATAGC CTGGGTGGCT TAAAC 1965

40 (2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 503 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

50 AGTGATCCCC TTGCCTCGGC CTCCCAAAAT GCTGGAATTG TAAGCGTGGG CCTCTGCACC 60  
 CGGCCTGGTC CGCAATTAA AAACGCACAG CCACCATTC GTYTCCAGAA AGCACCAGCA 120  
 55 TGCTTTTGGG AGAACCAGCC TCCTCCATGG AGGAAAGCTT GGGATCTGCC TTCCCACCTG 180  
 GGGAGGAGAG GGATCTGTGG AAAATCCTTC TGACCGACTT CCCCTCAGTG CCGATCCAT 240  
 ACTCAATAGT AGAAAAAGTA AGAAATATAC AAAGATAGCA GATACACGGA GACAGTTCCC 300  
 60 CAAATAGCTG AGCGAWTAGC GCAGAAGCAA TATTGAAGAC CTAATAGCTG AGACATTTCC 360

371

AGAACTGATA AAGTGCATCC AGCCACAGAT CAAGCAGCCC AGAAAATTCC AGGCAGCATC 420  
 AACAAATAAA TAGCCCCACA TGCACCCCTG AAAATGCAGA AGACCAACA AAAAAGTCCG 480  
 5 GTCAACAGCC AGAGTTAAAG AGG 503

10

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 1133 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

20

GGCAGGCTT GGAATGAACC CCGTGGATA AGGGGACTA TTAGATAGAA TAAACATCAA 60  
 TAAATGCTTG ATGAATAAAC GCTAATCCTA CCTTCCCAGC CTGACACCTC CCAGTGGACA 120  
 25 CCACACTTCA CTTGAAGCCT TAGAACCCTT TCCCACCAT CCTTCCAGCC CTGGCTTCAT 180  
 GTTGCCATTT CTCACCCCCA GAACAGGCCG CCGGCTGAA GAACTACAA GACCAAGAGA 240  
 AACAAACAGAA AGTGGAGTTT CGTAAAAGGA TGGAGAAGGA GGTGTCAGAT TTCATTCAAG 300  
 30 ACAGTGGGCA GATCAAGAAA AAGTTTCAGC CAATGAACAA GATCGAGAGG AGCATACTAC 360  
 ATGATGTGGT GGAAGTGGCT GGCTGACAT CCTTCTCCTT TGGGGAAGAT GATGACTGTC 420  
 35 GCTATGTCAT GATCTTCAAA AAGGAGTTTG CACCCCTCAGA TGAAGAGCTA GACTCTTACC 480  
 GTCGTGGAGA GGAATGGGAC CCCCAGAAGG CTGAGGAGAA GCGGAACNTG AAGGAGCTGG 540  
 CCCAGAGGCA ANGAGGAGGA GGCAGCCCAG CAGGGGCTG TGGTGGTGAG CCTTGGCAGC 600  
 40 GACTACAAGG ACAAGTACAG CCACCTCATC GGCAAGGGAG CAGCCAAAGA CGCAGCCCAC 660  
 ATGCTACAGG CCAATAAGAC CTACGGCTGT KTGCCCGTGG CCAATAAGAG GGACACACGC 720  
 45 TCCATTGAAG AGGCTATGAA TGAGATCAGA GCCAAGAAGC GTCTGCGGCA GAGTGGGGAA 780  
 GAGTTGCCGC CAACCTCCTA GGCGCCCCGC CCAGCTCCCT TTGACCCCTG GGGCAGGGCA 840  
 GGGGSCAGGG AGAGACAAGG CTGCTGCTAT TAGAGCCCAT CCTGGAGCCC CACCTCTGAA 900  
 50 CCACCTCCTA CCAGCTGTCC CTCAGGCTGG GGGAAAACAG GTGTTTGATT TGTACCGTT 960  
 GGAGCTTGA TATGTCCGTG GCAATGTGTG GTGTGTGTGA GAGTGTGAAT GCACAGGTGG 1020  
 55 GTATTTAATC TGTATTATTC CCGTCTCTTG GAATTTTCTT CCCATGGGGC TGGGTACTTT 1080  
 TACATTCAAT AAATACTGTT TAACCCAAAA AAAAAAAAAA AAAAGAAAGA AGN 1133

60

## (2) INFORMATION FOR SEQ ID NO: 119:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1101 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

GGGCACAGCT GAAGTGCAG ACCTCCCCAG GGGATGGCTC CTCTCCCCCA GGAGCCCCGA 60  
GGCAGGGGAG GCAGAAAGCC TGGGCTCTGG GGGGTGGCCT GCGGACAGCT GTGCTGTGGG 120  
CCGGGGGCTG GGCCTGTCCC ACAGGNCCT GGAGCTCGTG GTTCTGAGCA GCCAGCTGGG 180  
TGCTGTCTGG GCATAGCTGG GAGGCACAGC GGCTGCCATG TGGGACTGGG ACTGCACTGC 240  
TCCCTGGTCT TGGCCTCTGT GGCTCAGCCT TGCTCTGGTC TGCTGAGTG CAGGGGCCAA 300  
GGGCACAGG GCCAGTGAGG CCGGCCACGC TCGGGCCCTC ACCTGTGAGA TGGGTTCGGA 360  
ATTTKACACA GCCTANGGCT TGGTTCTTGG TKGNGAMCG TGGACTYCTK AGAACGGGAG 420  
TGCTGGTCCT .GAAAGGGCTG GTTGAGACG AGCTGCTTTT CTCGCTGTTT TTCTCTTAGG 480  
AGATTAAACA AAAACAGAAA GCACAAGAGC AACTCAGTAG CAGACCCCAG ACTCTCCCCT 540  
TGCCAGACGT GGTTCAGAC GGGGAGACGC ACCTCGTCCA GAACGGGAT CAGCTGCTCA 600  
ACGGGCATGC GCCGGGGGCC GTCCAAACC TCGCAGGGCT CCAGCAGGCC AACCGGCACC 660  
ACCGACTCCT GGTGGGGGCC CTGGCGAAGT TGTGTGTGAT AGTTGGGTTT GCAGCCTTTG 720  
CTTACACGGT CAAGTACGTG CTGAGGAGCA TCGCGCAGGA GTGAGGCCCA GGCGCCGAGA 780  
CCCCAGGCGC CACTGAGGGC ACCCGGCACC AGAGCGTGAC CTCGGCAGGC TGGACACACT 840  
GCCCAGCACA GGCAGACCCA CCAGGCTCCT AGGTTTAGCT TTTAAAAACC TGAAAGGGGA 900  
AGCAAAAACC AAAATGTGTG ACTGGGCTTT GGAGGAGACT GGAGCCTCAG CCCTGTCTTG 960  
GCCACGGGCC GCTGGGGCTG GTGTGGGTGG GCCTTGTGTG CTGGATTTGT AGCTTATCTT 1020  
CCGTGTGTG TTTGGACCTG TTTTAGTAAA CCCGTTTTTC ATTTTAAAAA AAAAAAAAAA 1080  
AAACTTTGGG GGGGGGCCCC N 1101

## (2) INFORMATION FOR SEQ ID NO: 120:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 232 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

AGCTTCTCTG TCCAGTCTTG AACTCTGGGS TCTCTTGGAA CTTTCCTCAG CCGTCTCAGC 60  
 5 CTGAATATTC CTTCCATGGA TTCCACTCAA CCAGACTTTG GATCTGTGCC TACTTATCA 120  
 ACCTTATCTT TGCAATATGT TCGGGCCAC CTTCCACTCC TTGGTTCTTG TTCCTCCTG 180  
 GCCTAACTTG TCCCTTCTCC ACTTCACATC CCGGTGGGA CAGCATTCCT CTTTCCTCCC 240  
 10 AACCTCCCTC CGTCTCRAA AAAAAAAAAA AAAAAAAAAA TT 282

15

## (2) INFORMATION FOR SEQ ID NO: 121:

## (i) SEQUENCE CHARACTERISTICS:

20

- (A) LENGTH: 2535 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

TAAGGGGGTG TGTGCTCACC TCCTCCTGAC CCTTAACACT CCTGTCTGCG CCAGACCAAC 60  
 AGAGAGAGCT GTCCTTGAGA CCGCGGAGAG AAGCAGCTGC CGAAAGCTGC AGCCTTTCCG 120  
 30 CACTCTGAGA CATGATCTT CCTCCTGCCA GGGGAGAGCC ACCCAGAGCC CATGTCCAGC 180  
 CCCACTTCCC TCAGCCCCCA GGGTTCTCTT CTGGCCCTC TGAGGATTCG CTAGGGCTGC 240  
 CCGGCAGAGG GGTTCCTCCA AGCTCTGTTT TGAAGCCTGC AATGTGGAAA AGTGAGAAGT 300  
 35 CAGAGGGAAC AGGACAGGTG CAGCCGGGCT CTGAGGCCAC ACCTCAGACC TCGCTGTTCG 360  
 CCAACATCCC CTGAGCAGTG TGAGCTCATC TCACCAGATG AGAAGAGGCC CTGTGCATTT 420  
 40 YTTTGTGTTG TTTGTGCTG TTTTCCCCCA CCCATCCAGT TCTCCTCAGC AAAGCAAATT 480  
 CCTTAACACC TTTGGTGGAG AATTTCTTAC CCAGACTTGG GGCTGTGATG CCCTTCAGTG 540  
 CGTGGTGAGT GCAGCGTGTG TCGGTGTGCC TGTGTGTGAA CCTGGGGGCC ATCCTGGTGG 600  
 45 CCTGGGAGCG TGAGGAGAGG CCCCCTGTGT GCTGGGTGAG TGGTGGGTGT GGGGTCAATG 660  
 CAGTGAGGCT CTCTGGGTGA GGCTCCCAAC CTGGCAGTCC CCAGCCTCCC AGCATCTGTG 720  
 50 AGCGTCTGTT GGAATTACA GAAGAGCCTC ATCCGTCTCTG CCCCTCACTC TGCCCTGGAA 780  
 TCAACATCTT CCGAGTCCTT CTTGGGGGAA ATAGCAGAGC CCCACTTAAC TCCATAAACT 840  
 GCTTCCCATT CCGCAGCCCA GTTCTGATTG TTGAGGTGTC GCGTCTGTCC AGGTCCCCCA 900  
 55 GTCCCTCTTT TCTCTGTCC TCTCTCTGTG CTTCACTTCC CCACTCCAGC CCGGGCTCAG 960  
 TTCAGGGAAA TGCTGTTCCTA YATCAGCCCT CTGCTCTCTG AGGCAGCCGC GCCTCTGACT 1020  
 60 CGGAGCTACT TGAAACTTCT GCTCTTGCTA GGATTGGAGT CTACCTATCT CTTCCATTTG 1080

	TCCCAGCTGG AGTTCTGGAA CTTTCTCTCT CCGGCTGGGG GTGGGGGTTG TTAAGGATGC	1140
5	TGGGGGGCCT GGGGAAGGAA GGAGTTCAGA GGAGGGTGT CCCCCTGTCT CTTGATGTCA	1200
	CCCTCCGCTC CTGGGACACG TGCTCTCTCT GTCTCTGGGT CTTCTGGGTTG TGCACGTTTG	1260
	TGTGTCTTGG TAAATATGTT TTAGGAAGAA AGCAAAAGGG ACTGAACTAG CCTTTGGTAG	1320
10	GATTGCAGGG GTCCAGCCCT GCCTGTTTCC GAAGCCCCCA CACTGCTTTT CCCCCACTG	1380
	AGACTGGTCC CCTCAAAAGG TAGACAAAC AGCACTCCC TGTGGAGTTG AAGGGGGCC	1440
15	TCAAAGTGGC TTTTGTGTAG ACAAGGTTAA GGTTCCTCA TGAGCAATTT TGCAGATCGG	1500
	TCCTTCTCA CCTCCTTGAT TTGTGACCTT GACCAAGGGG CCGTGCCTCC AGCTCTCTCA	1560
	GTGCCCTCTC CTGATGCTT CGCTGCTTCC TGGCCCTACT CCGCTGCTTT AGGTAAGTTG	1620
20	GGGAATTAGG GCCATGCTGG AAGAAGCTTA ACCGTGTGTT CAAGAAGGG TTTTGTGCTT	1680
	GCTTGGTCTT GGAAGTCCCC TTGGCTGCCC CAGGCTCTCT TGGCCCTGG GTGTGGGGG	1740
25	AGGTGGATGT CAGATCTGGT AGGTTCAGC AGAGAAATA AATGTGCTTT GAGAGACCC	1800
	TCAGAGAGGG TCCAAGGGTG ATGGAGAAGG AAGCATGGCC TGGGAGTTG GATGGGAGG	1860
	GTGGTGGGTG CCGGCATCTT GACTGCCCCC TGTGTCTCCA CAGTGGGGG GTGTGACCC	1920
30	CYTTCCTCTC CAGCCCGCCT GCCTTCAGCC TTCTCTAGC TTCACTGCTT TCTACTTCA	1980
	CTTTGGAGGG GGTGGGGTCC GTTGGCATCA ACACGGGAC CCTGTGTTT ACCAAAGCCC	2040
35	GAGCCCTCAG CCCCTGGGGA GAACAAATGG CTGAGCTTTG ATAGCTGGG TCTTCGAGG	2100
	GCTGGGGGCT GCGGGCAGTC CCAGGGGAGA GACACCAAG AAGGAGATC AGAATCCCTG	2160
	AGGAAGTTCC CAGCAGAGCA AACTGCTTTC CAGCCTGAG CCTGCTTAAA CTGTGTGATG	2220
40	TGCAATAACT GAGCTTAGAG TTAGGAATTG TGTTCAAATG CTGGGATTC CGTCTGTAGA	2280
	TTTAACTGCT GAAATTGTAT CTCTCAGTAA TTTTAGATGT CTTTAAAAA ATGAAAAAC	2340
45	AAAGTGTTAG ACTGTGTGG TGTGGTTGA TGGGCACTCA AGATGCTCTT GATCTATCCA	2400
	GCCCTGCCTT TCCCCTGCGC CCCCATCCTC TCAGTCCCG CCGTGGCTTC ACTTGGGGAC	2460
	CCTGCCTCGT GTGCTCTTTA TCTGCCTATT ACTCAGCTTA AGGAACAGG TACATCCAC	2520
50	ACATGCATAA AGGAAATCAA ATGTTATTTT TAAGAAAATG GAAATAAA ACTTTATAAA	2580
	CACCAAAAAA AAAAAAAAAA ACCGNGGGGG GGGGGCGGTA ACCCATTTGG CTTAA	2635

55

(2) INFORMATION FOR SEQ ID NO: 122:

(1) SEQUENCE CHARACTERISTICS:

60

(A) LENGTH: 994 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

	GAATTCGGCA GAGGTTGGG GAAGATAGGG AATAAGGAAG CACAGGAGTA GGGGACAAGG	60
10	AAGCACAGGA GTAGGGGAGA TATACAGCGG TCAGGATAAG GGGGAAAGGG CGGTGGTTGC	120
	SCAAGAGGTG AAACAAGATG TGAGAGACAA GGGGTAGCGA AGAAATGGGG CAGCGGTTAG	180
	GTTGAGAAGC GCATAGACCG TGGCGGACGG GCAATGCCAG GGGCACAGAA AGGAAGTGAG	240
15	GGGTGGGCTA TTTAARGGA GATGCTCCTT CAGCCCTCTT YTTTCTGCG TAGTTCTCCT	300
	CCTCCAGGCC GCGCGCGGAT ATGTCGTCCG GAAACCAGCC CAGTCTAGGC TGGATGATGA	360
20	CCCACTTCCT TCTACGCTGC TCAAAGACTA CCAGAATGTC CCTGGAATG AGAAGGTTGA	420
	TGATGTCTG AAAAGACTCT TGTCTTTGGA AATGGCCAC AACAAGGAGA TGCTAAAAAT	480
	CAAGCAAGAA CAGTTTATGA AGAAGATTGT TGCAAPCCCA GAGGACACCA GATCCCTGGA	540
25	GGCTCGAATT ATTGCCCTGT CTGTCAAGAT CCGCAGTTAT GAAGAACAAT TGGAGAAACA	600
	TCGAAGGAC AAAGCCCA CAACGCTATCT GCTAATGAGC ATTGACCAGA GGAAAAAGAT	660
30	GCTCAAAAAC CTCCGTAACA CCAACTATGA TGTCTTTGAG AAGATATGCT GGGGGCTGGG	720
	AATTGAGTAC ACCTTCCCCC CTCTGTATTA CCGAAGAGCC CACCGCCGAT TCGTGACCAA	780
	GAAGGCTCTG TGCATTGGG TTTTCCAGGA GACTCAAAAG CTGAAGAAGC GAAGAAGAGC	840
35	CTTAAAGGCT GCAGCAGCAG CCAAAAAACA AGCAAAGCGG AGGAACCCAG ACAGCCCTGC	900
	CAAAGCCATA CCAAAGACAC TCAAAGACAG CCAATAAATT CTGTTCAATC ATTTAAAAAA	960
40	AAAAAAAAAA AAAAAAAAAA AAAAAGGGGA GGGG	994

45 (2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1542 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

55	GGCAGAGCCA CCTCGGCCCC GGGCTCCGAA GCGGCTCGGG GCGGCCCTTT CGGTCAACAT	60
	CGTAGTCCAC CCCCTCCCCA TCCCAGCCCC CCGGGGATTC AGGCTGGCCA GCGCCAGCC	120
	AGGGAGCCCG CCGGGAAGCG CGATGGGGC CCCAGCCGCC TCGCTCCTGC TCCTGCTCCT	180
60	GCTGTTCGCC TGCTGCTGGG GCGCCGCGCG GGCACACCTC TCCCAGGACG ACAGCCAGCC	240

	CTGGACATCT GATGAAACAG TGGTGGCTGG TGGCACCGTG GTGCTCAAGT GCCAAGTGAA	300
5	AGATCAGGAG GACTCATCCC TGCAATGGTC TTAACCCCTGC TCAGCAGACT CTCTACTTTG	360
	GGGAGAAGAG AGCCCTTCGA GATAATCGAA TTCAGCTGGT TAMCTCTACG CCCCACGAGC	420
	TCAGCATCAG CATCAGCAAT GTGGCCCTGG CAGACGAGGG CGAGTACACC TGCTCAATCT	480
10	TCATATGCC TGTCCGAAC TCCCAAGTCCC TCCTCACTGT GCTAGGAATT CCACAGAAGC	540
	CCATCATCAC TGGTATAAA TCTTCATTAC GGGAAAAGA CACAGCCACC CTAAACTGTC	600
	AGTCTTCTGG GAGCAAGCCT GCAGCCCCGG TCACCTGGAG AAAGGGTGAC CAAGAACTCC	660
15	ACGGAGAACC AACCCGCATA CAGGAAGATC CCAATGGTAA AACCTTCACT GTCAGCAGCT	720
	CGGTGACATT CCAGGTTACC CGGGAGGATG ATGGGGCGAG CATGGTGTGC TCTGTGAACC	780
20	ATGAATCTCT AAAGGGAGCT GACAGATCCA CCTCTCAAGC CATGAAGTT TTATACACAC	840
	CAACTGCGAT GATTAGGCCA GACCCCTCCC ATCCTCGTGA GGGCCAGAAG CTGTTGCTAC	900
	ACTGTGAGGG TCGCGGCAAT CCAGTCCCCC AGCAGTACCT ATGGGAGAAG GAGGGCAGTG	960
25	TGCCACCCCT GAAGATGACC CAGGAGAGTG CCCTGATCTT CCCTTTCCTC AACAAGAGTG	1020
	ACAGTGGCAC CTACGGCTGC ACAGCCACCA GCAACATGGG CAGCTACAAG GCCTACTACA	1080
30	CCCTCAATGT TAATGACCCC AGTCCGGTGC CCTCCTCCTC CAGCACCTAC CACGCCATCA	1140
	TGGGTGGGAT CGTGGCTTTC ATTGTCTTCC TGCTGCTCAT CATGCTCATC TTCCTTGGCC	1200
	ACTACTTGAT CCGGCACAAA GGAACCTACC TGACACATGA GGCAAAAGGC TCCGACGATG	1260
35	CTCCAGACGC GGACACGGCC ATCATCAATG CAGAAGGCGG GCAGTCAGGA GGGGACGACA	1320
	AGAAGGAATA TTTTATCTAG AGGCGCCTGC CCACTTCCTG CGCCCCCAG GCCCCTGTGG	1380
40	GGACTTGCTG GGGCCCTCAC CAACCCGGAC TTGTACAGAG CAACCGCAGG GCGCGSCCCT	1440
	CCCGNTTGTT CCCCAGCCCA CCCACCCCT TGTACAGAA TGTATKGTTC GGGGTGCGGT	1500
45	TTTGTWATTG GTTTNGGATN GGGGAAGGGA GGGANGGCGG GG	1542

## (2) INFORMATION FOR SEQ ID NO: 124:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1390 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

CAAGCTCTAA TACGACTCAC TATAGGGAAA GCTGTACGC CTGCAGGTAC CGGTCCGGAA 60



TTCCCGGGTC CACCCACGGC TCGGGGCTC AGGGTGGACG CATGGTTCTG CACTGAGGCC 120  
 CTCGTCATGG TGGCGCTGT GTGGTACTTG GTAGCGGCGG CTCTGCTAGT CGGCTTTATC 130  
 5 CTCTTCTGTA CTCCGAGCGG GGGCGGGCGG GCATCAGCGG GCCAAGAGGC ACTGCACAA 240  
 GAGGAGCTGG CAGGAGCAGG CCGGGTGGCC CAGCCTGGGC CCCTGGAGCC TGAGGAGCCG 300  
 10 AGAGCTGGAG GCAGGCCTCG GCGCCGGAGG GACCTGGGCA GCGGCTACA GCGCCAGCGT 360  
 CGAGCCACAG GGTGGGCTG GGCAGAAGCA GATGAGAAGC AGGAGGAAGC TGTCTCTCTA 420  
 GCGCAGGAGG AGGAAGGTGT CGAGAAGCCA GCGGAAATC ACCTGTGGGG GAAATTTGA 480  
 15 GCTAAGAAAC TCGGAANNT GGAGGAGAAA CAGCGCGAA AGGCCAGCK TGAGGCAGAG 540  
 GAGGCTGAAC GTGARGWGG GAAACGACTC GAGTCCGAGC CGGAATGAGT GGAAGAAGGA 600  
 GGAGGAGCGG CTTCGCTTG AGGAGGAGCA GAGGAGGAG GAGGAGAGGA AGGCGCGCA 660  
 20 GGAGCAGGCC CAGCGGGAGC ATGAGGAGTA CCTGAACTG AAGGAGGCTT TTGTGGTGA 720  
 GGAGGAAGGC GTAGGAGAGA CCATGACTGA GGAACAGTCC CAGAGCTTCC TGACAGAGTT 780  
 25 CATCAACTAC ATCAAGCAGT CCAAGTTGT GCTCTTGGAA GACCTGGCTT CCCAGGTGGG 840  
 CCTACGCACT CAGGACACCA TAAATCGCAT CCAGGACCTG CTGGCTGAGG GCACTATAAC 900  
 AGGTGTGATT GACGACCGG GCAAGTTCAT CTACATAACC CCAGAGGAAC TGGCCGCCGT 960  
 30 GGCCAACCTC ATCCGACAGC GGGCGCGGCT GTCCATGCC CAGCTTGCCC AAGCCAGCAA 1020  
 CTCCTCATC GCCTGGGGCC GGGAGTCCCC TGCCCAAGCC CCAGCCTGAC CCCAGTCTT 1080  
 35 CCCTCTTGA CTCAGAGTTG GTGTGGCTTA CCTGGCTATA CATCTTCATC CCTCCCCACC 1140  
 ATCCTGGGGA AGTCATGCTG TGGCAGGCA GTTATAGATT AAAGGCTGT GAGTACTGCT 1200  
 GAGCTTGGTG TGGCTTGGTG TGGCAGAAGG CCTGGCTAG GATCCTAGAT AAGCAGGTGA 1260  
 40 AATTTAGGCT TCAGATATA TCCGAGAGT GGGGAGGCTC CCTTGAAGC TGGTGAAGTC 1320  
 CTGTTCTTAT TATGAATCCA TTCAATCAAG AAAATAGCCT GTTGCAAAAA AAAAAAAAAA 1380  
 45 AAAAACTCGA 1390

50 (2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1288 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

60 GCGCGCGCGG TGAAAGGCGG ATTGATGCGG CTTGGGCGG CTTGGGAGCG CGCGGAGCA 60

5 GACGCTGACC ACGTTCCTCT CCTCGGTCTC CTCCGCTCC AGCTCCGGCG TGGCCGGCAG 120  
 CCGGGAGCCA TCGACACCCA GGGCCCCGCC GCTCCCGCC AGCGGCTCCG CGGCCTCCTG 180  
 CTGCTCCTGC TGCTGCAGCT GCGCGCGCCG TCGAGCGCCT CTGAGATCCC CAAGGGGAAG 240  
 CAAAAGGGCG ATCCGGCAGA GGGAGGTGGT GGACCTGTAT AATGGAATGT GCTTACAAGG 300  
 10 GCCAGCAGGA GTGCCTGGTC GAGACGGGAG CCTCGGGGCC AATGGCATTG CGGGTACACC 360  
 TGGGATCCCA GGTCCGGATG GATTCAAAGG AGAAAAGGGG GAATGTCTGA GGGAAAGCTT 420  
 TGAGGAGTCC TGGACACCCA ACTACAAGCA GTGTTTCATG AGTTCATTGA ATTATGGCAT 480  
 15 AGATCTTGGG AAAATTGCGG AGTGATACAT TACAAAGATG CGTTCAAATA GTGCTCTAAG 540  
 AGTTTTGTTC AGTGGCTCAC TTCGGCTAAA ATGCAGAAAT GCATGCTGTC AGCGTTGGTA 600  
 20 TTTCACATTG AATGGAGCTG AATGTTTCAGG ACCTCTTCCC ATTGAAGCTA TAATTTATTT 660  
 GGACCAAGGA AGCCCTGAAA TGAATTCAAC AATTAATATT CATCGCACTT CTTCTGTGGA 720  
 AGGACTTTGT GAAGGAATTG GTGCTGGATT AGTGCATGTT GCTATCTGGG TTGGCACTTG 780  
 25 TTCAGATTAC CCAAAGGAG ATGCTTCTAC TGGATGGAAT TCAGTTTCTC GCATCATTAT 840  
 TGAAGAATA CCAAATAAAA TGCTTTAATT TTCATTGCT ACCTCTTTTT TTATTATGCC 900  
 30 TTGGAATGGT TCCTTAAAT GACATTTTAA ATAAGTTTAT GTATACATCT GAATGAAAAG 960  
 CAAAGCTAAA TATGTTTACA GACCAAAGTG TGATTTTACA TGTTTTTAAA TCTAGCATTA 1020  
 35 TTCATTTTGC TTCAATCAAA AGTGGTTTCA ATATTTTTTT TAGTTGGTTA GAATACTTTC 1080  
 TTCATAGTCA CATCTCTCA ACCTATAATT TGGGAATATT GTTGTGGTCT TTGTTTTTTT 1140  
 CTCTTAGTAT AGCATTTTTA AAAAAATATA AAAGCTACCA ATCTTTGTAC AATTTGTAAA 1200  
 40 TGTTAAGAAT TTTTTTTATA TCTGTAAAT AAAATTATT TCCMACAACC TTAACAAAAA 1260  
 AAAAAAAAAA AAAAAAAAAA AAAAAA 1288

45

(2) INFORMATION FOR SEQ ID NO: 126:

50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1517 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

60 AGTGGCTTAA AGGCATCGTT TTAGGGATTA CTGGGAAGTA TCTTCAAAGT AATACATGAG 60  
 AAACATTCTT TCCTAAATCC TTTATTATAT TGAATATCCT ATTAATTGGT TTTCAGACGT 120

	TAAATTAACC ATGTATTCCCT GCAATAAATG TCACTTGTNT CTGTATATA ATCTTTTTTA	180
	TATATTACCG GATTGATCA TTAGTATTTT GTTGAGGATT TTTGTGTCTA TATCATAAG	240
5	AGATGCTGGT CTGCAGTTTT CTTTTTTTGT GATAATCTGG TTTTGTATC AGTAATACAG	300
	GCCCCATGAA ACCAGTTGGG AAGTGTTCAC CTCTCTGTA TTTTTTCAAG AGTTTGTGAA	360
10	GAATTGCTAT TAATTCTTTA AATGTTTGGT AGAATCTACC ATTGAAATCA TGTGCTCTGG	420
	GCTTTTTTTT GAGGGAAGTG TTCTGATAAC TAATTCACTA TCTACTTTTT ATAGCTCTGT	480
	TCACATTTTG CTCTTCTCG AGTTAGTTTT GGTAATTTGT GTATCTCTAG GATTTTGTCC	540
15	ATTTCAATTA TCTCAATTTG TGGCATAAAT TAAACTAAAT TTGGCCTGAG CCTACCTGTA	600
	TATCTTGAGT CCCTCTGTAA GGAAGTGTAG CCTAAGTTGT ACATAAACAA ACTGAAATCC	660
20	TAAATTAGGA ATGTAGTTTT TGTAAACAGT CCTGAGTCTC AGGCAGTCAC AGCAG/CAAG	720
	TCTGTCAATT GCAGGCTGCT AACTAAGCAG CCCATGSTCA AATGAGGCAA AAACCTTTGC	780
	TTTTAACACA TAGTATAGCT TTGTAATCCT TTTCTTGAC ACTCGGGTAA TTTCTTCTT	840
25	TTTCATTCCC KGWATTTTCC AKGAATATGA RTCTYCCTTT TTTCCCTCC TGTCACTCTA	900
	GCTAATGGTT TGTCAATTTT GTTGATCTTT TGAARAACAA ACCTTTGGTT CCACTTTCTT	960
30	GTTCATATG CTGARTATTC TCATAATTGG AGTGGAAAGC TGATCTTTGA TTACTTATTT	1020
	TACTTAGGGC TGAGGAGTTC ATGGACTTCG CAAAACCTCC TTGAATCTAA ATTGCATCTT	1080
	CTTTCCTGGT TTCTGGGCTG AAACATGTTT TTTCCCATCT WANAWACCCT TGGTCTTTTC	1140
35	ATKGGCGATT AAGACTAGAG AAAGTTCTAG ATMCCTTGTC CTTTTATGCT GTCATTTTGT	1200
	TTAAAGGCTT TCTATGTAGT AAAACTATCT ATATAGACAA AATAGAGCCT TGAGTTGTGG	1260
40	TCTTGAATTT GATCAACATG ATTTACCACA TTCTGTACTG GATATTTCTT CACCTGCTGC	1320
	TACTGTAAAC CATTTTATTC TTGGATCTTC TGTAGAGTAT ATTATCACAG GTACTTTTTA	1380
	CAGGGGTGTC TAATCTTTTG GCTTCCCTGG GCACATTGAA AGAAGAAGAA TTGTCTTGGG	1440
45	CCACACATCA AATACGCTAA CACTAATAAT AGTTGATGAG CTAACAAAAA AAAAAAAG	1500
	GCAAAAAAGN CCCAAAA	1517

50

(2) INFORMATION FOR SEQ ID NO: 127:

55

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1073 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

5 TGAATCTATT CTTTGAACAT TCTACAACAA GAATTACATT ATACTGTTAT ACCAGAGTAC 60  
 TTCTGCAGTG TGAAATAGAT TGGTTTGGAA AATGAACCTG GCTTTGCTAT AAATTACATT 120  
 CACRGGCCTT TTTGCAAAATG TGTAACCTGC CTATCAAAGT AGTTTGTAGG GCAAATGCAG 180  
 AATATATGTC TCCATCTGGT AAAGTACCTT WTAATCATGT GGGAAATCRA GTAGTATCAG 240  
 10 AACTTGGTCC AATAGTCCAA TTTGTTAAAG CCAAGGGCCA TTCTCTTAGT GATGGGCTGG 300  
 AGGAAGTCCA AAAAGCAGAA ATGAAAGCTT ACATGGAATT AGTCACAAAT ATGCTGTTGA 360  
 CTGCAGAGCT GTATCTTCAG TGGTGTGATG AAGCTACAGT AGGGRMGATC ACTCATGMTA 420  
 15 GGTATGGWTC TCCTTACCCT TGGCCTCTGW WTCATATTTT GGCCTATCAA AAACAGTGGG 480  
 AACTCAAACG TAAGNTGAAA GCTATTGGAT CGGGAAGAA GACTCTGGAC CAGGTCTTAG 540  
 20 AGGATGTAGA CCAAGTGCTGT CAAGCTCTCT CTCAAAGACT GGAACACAA CCGTATTTCT 600  
 TCAATAGCA GCCTACTGAA CTGACGCAC TGGTATTTGG CCATCTATAC ACCATTCTTA 660  
 CCAACAAATT GACAAATGAT GAACTTTCTG AGAAGGTGAA AACTATAGC AACCTCCTTG 720  
 25 CTTTCTGTAG GAGAATTGAA CAGCACTATT TTGAAGATCG TGCTAAAGGC AGGCTGTCAT 780  
 AGAGTTATGT GTTAGTCTCA GGAGTCTTAA CTTTGAAT ATGTTTTACT TGAATGTTAC 840  
 30 ATTAGATAAT GGTGTCAGAA TTTTAAACC AAATTACTGC TTTTGAAAC CTCAAATTAT 900  
 ATAATGTATC TTATGTATGT GCTTTATATT GTTATTTGTG TATACATTAA AATAATTCTG 960  
 AATTATTTAA TCTGATATGT TGTATTCTGT ATCTTGAAT TTTTGTTCCT TTGAAACATG 1020  
 35 CATGCATTAA AAAATAAACC TTAAACAAC GTAAAAAAA AAAAAAAA CTC 1073

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 300 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

50 CAACCCCTGC CTTTTTTTTG TTTTCCATTT GCTTGGTAGA TCTTCCTCCA TCCCTTTATT 60  
 TTGAGCCTAT GTGTCTCTCT GCCCGTGAGA TGAGTCTCCT GAATACAGCA CACTTACTGG 120  
 55 TCTTGACTCT GTATCCAATT TGCCAGTCTG TGTCTTTCAT TTGGAGCAAT TAGCCCATTT 180  
 ACATTTAAGG TKAATATTGT TATGTGTGAA TTTRATCTR TCATTATGWT GTTAGCTGGT 240  
 60 TATTTTGCTT GTTAGTTGAT GCAGTTTCTT CCNGGCATCA ATGGTCTTTA CAANTTGGCA 300

## (2) INFORMATION FOR SEQ ID NO: 129:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1275 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

5	GGCAGAGCCT GTCCCTGCTG CCCCTGCAAA AAAAACCCCC TCTGGTGTGA GCAGGATGGT	60
15	TGGAGGTTAT GTGAGCTCCT TCTCCTTTCC TCCAGTTTCC TCTTCCCTTC TCCTCCCTGC	120
	CTCTTTTGCT TTCCCTTTC TTCTGGTAC CCCCTGCCCA TTCCTGTATT TTCTCCCATC	180
20	GCCATTCTCC CCTCTCCAC TGTCCCTAAC CCGTTCAAAC TCTTCCCTCT TAAATGGTTG	240
	AGATTTTCTC TCACCAAGCA CACCCAGTA TTAATTAAAC TAGCTGCAAA CAGGTAGCAA	300
25	GTGGTCTACC ATGACAGATG GGTTTTGTGT GTGTGTGTGT GTGTGTAATT GTAATAAAC	360
	ATATTGARTC ACTCAATAAA CACAGAGTGT CTAATACATG TATCARGCAC TATCATAGAT	420
	GCTAATTAAC GAAACTGAAA TGGCCAGGCC CTCACAGTGG CTCATGCCTA TAATCCCAGC	480
30	ACTTTGGGAG GATGAGGCAG GAGGATCACT TGAGGCCGGG AGTTCAAGAC CAGCCTGGGC	540
	AACATAGTAA GACTCCATCT CTACAAAAA AAAATTTTTT TTATTATACT TTAAGTTTTG	600
35	GGTTACATGT GCAGAACGTG TAGTTTTGTT ACATAGGTAT ATACGTGCCC TGGTAGTTTG	660
	CTGCACCCAT CAACCCATCA CCTACATTAG GTATTTCTCC TAATGTTACC CCTCTCCTAG	720
	CCCCCACCC CGTGACAGGC CCTGGTGTGT GATGTTCCCC TCCCTGTGTC CATGTGTTCT	780
40	CATTGGTCAA CTCTCACCTA TGGAGTGAGA ACATGTGGTA TTTGGTTTTT TGATCTGTG	840
	ATAGCTTGCT GAGAATGTTG GTTCCAGCT TTATCCACGT CCCTGCAAAG GGCATAAACT	900
45	CATCCCTTTT TATGGCTGCA TAGTGTCCA TGGGTATAC GTGCCACATT TTCTTAATCT	960
	ATCATTGATG GACAAGTTTT GCTATTGTGA ATAGTCCCAQ AATAAACATA CGTGTGGGTG	1020
	TGTCTTTATA GCAGCATGAT TTATAATCCT TTGGGTATAT ACCCAGTAAT GGGATCACTG	1080
50	AGTCAAATGG TATTTCTCGT TCTAGATCCG TAAGGAATTG CCACACTGTC TTCCACAATG	1140
	TTTGAACATA TATCACTCC CACCAACAGT GTAAAAGTGT TTCTATTTTT CCACAACCTC	1200
55	TCCAACATCT GTTATTTCT GACTTTTTAA TGAACGTGAT TCTAACTGGC GTGAGATGGT	1260
	ATCTCATTGT GGTTC	1275

## (2) INFORMATION FOR SEQ ID NO: 130:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 472 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

CNGAACCCTC GTGAACCTC CCCGGGTAA AAAGCCCCCT CTAATGGGG GGAACGCTC 60  
 ACACGTTATA AAAAGCACT AGAATGTTTT GAAAGCGAGA AACAACAGCT GTGTAGGTA 120  
 GCTAGCACTT AGTGTGTAC AGAAGACAGA TATTTGTGCA TTTGTGCATT TTCTAAGTTT 180  
 GCTGCAATGA GCATGTATTA CTTTCATAGT TATAAACAC ATGCAAAATG CCCTTTTAAA 240  
 ATGAAAAAAA ATCCATGAGT GTAAGTGATA TATATGCTTT GGAAAGCCTG GGACGGTCAT 300  
 TGTCTACTCT CAATAGTATG TGTTCGCTT TGTCTTTTG AGACATTTTG TTTAATCTG 360  
 TTGATGACAA TAACCTGTG ATAATATAAC TTGATAACAA ATAAATGAC TTATGATTGA 420  
 AWMAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA NN 472

## (2) INFORMATION FOR SEQ ID NO: 131:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1950 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

ACCTCTCAGA ATCTTCTCTC AGCAACCTGA GTCTTCGCGG TTCTCAGAG CGCCTCAGTG 60  
 ACACCCCTGG ATCTTCCAG TCACCTTCCC TGGAAATCT GCTGTCCAGC TGCTCCCTGT 120  
 GCGGTGCGCTG TNAITCGCTG GTGTATGATG AGGAAATCAT GGCTGGCTGG GCACCTGATG 180  
 ACTCTAACCT CAACACAACC TGCCCTTCT GCGCTGCCC CTPTNTGCCC CTGCTCAGTG 240  
 TCCAGACCTT TGATTCCCGG CCCAGTGTCC CCAGCCCCAA ATCTGCTGGT GCCAGTGCCA 300  
 GCAAAGATGC TCCTGTCCCT GGTGTCCTG GGCCTGTGCT CAGTGACCGA AGCTCTGCCT 360  
 TGCTCTGGAT GAGCCCCAGC TCTGCAACGG GCACATGGGG GGAGCCTCCC GCGGGGTGA 420  
 GAGTGGGGCA TGGGCATACC TGAGCCCCCT GGTGCTGCGT AAGGAGCTGG AGTGGCTGGT 480  
 AGAGAACCAG GGCAGTGAGG TGCTGGCGTT GCCTGAAGTG CCTCTGCCC ACCCATCAT 540  
 CTTCTGGAAC CTTTGTGGT ATTTCCAACG GCTACGCTG CCCAGTATTC TACCAGGCCT 600  
 GGTGCTGCGC TCCTGTGATG GGCCTTCGMA CTCGAGGCC CCATCTCCTT GGCTAACCCC 660

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45

TGATCCAGCC TCTGTTTCAGG TACGGCTGCT GTGGGATGTA CTGACCCCTG ACCCCAATAG 720  
 CTGCCCACCT CTCTATGTGC TCTGGAGGGT CCACAGCCAG ATCCCCCAGC GGGTGGTATG 760  
 GCCAGGCGCT GTACCTGCAT CCCTTAGTTT GGCACGTGTG GAGTCAGTGC TGCGCCATGT 840  
 TGGACTCAAT GAAGTGCACA AGGCTGTGGG GCTCCTGCTG GAAACTCTAG GGGCCCCACC 900  
 CACTGGCCTG CACCTGCAGA GGGGAATCTA CCCTGAGATA TTATTCCTGA CAATGGCTGC 960  
 TCTGGGCAAG GACCACGTGG ACATAGTGGC CTTGATAAG AAGTACAAGT CTGCCTTTAA 1020  
 CAAGCTGGCC AGCAGCATGG GCAAGSAGGA GCTCAGGCAC CGGCGGGGCG AGATGCCCAC 1080  
 TCCCAAGGCC ATTGACTGCC GAAAATGTTT TGGAGCACCT CCAGAATGCT AGAGACCTTA 1140  
 AGCTTCCCTC TCCAGCCTAG GGTGGGAAG TGAGGAAGAA GGGATTCTAG AGTTAAACTG 1200  
 CTTCCCTGTT GCCTTCATGG AGTTGGGAAC AGGCTGGGAA GGATGCCCAG TCAAAGGCTC 1260  
 CAAGCGAGGA CAACAGGAAG AGGGATCCAC TGTTACCAA AGTCCTGATT CCCCCATCAC 1320  
 CAACCTACCC AGTTTGTTCG TGCTGATGTT GGGGGAGATC TGGGGGAGT TGGTACAGCT 1380  
 CTGTTCTTCC CTGTCTCTAT ACCGGGAAGT CCGCTCCAGG GTACCCACAG ATCTGCATG 1440  
 CCCTGGTCAT TTTAGAAGTT TTTGTTTTAA AAAACAAGT GAAAGATGCA GAGCTACTGA 1500  
 GCCTTTGCCC TGAATGGGAG GTAGGGATGT CATCTCCAC CAATAATGGT CCCTCTTCCC 1560  
 TGACGTTGCT GAAGGAGCCC AAGGCTCTCC ATGCCTTTCT ACCTAAGTGT TTGTATTTTA 1620  
 TTTTAAATTA TTTATTCTGG AGCCACAGCC CCCTTGCTTA TGAGGTTCTT ATGGAGAGTG 1680  
 AGAAAGGGAA GGGAAATAGG GCACCATGGT CCGGTGTTT GTAGTTCCTT CAAAGTCAGG 1740  
 CACTGGGAGC TAGAGGAGTC TCAAGCTCCC CTTAGGAAGA ACTGGTGCCC CCTCCAGTCC 1800  
 TAATTTTCT TGCTGCCCC GCTTGGGGA ATGCTCACC CACCCAGGTC CTGACCTGTG 1860  
 CAATAAGGAT TGTCCCTGC GAAGTTTTGT TGGATGTAAA TATAGTAAAA GCTGCTTCTG 1920  
 TCTTTTCAA AAAAAAAAAA AAAAAAACT 1950

(2) INFORMATION FOR SEQ ID NO: 132:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 990 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

TGGAAGATTT AAAATAGCTT TCATATTTCT CTTGAATATG AATATATAAG CTTGAATAAG 60

	CTTGAGTGGT TATTATATG AATTTTGGT TATTATTTCT ACCAATGCTT CTTATATTAA	120
	AGCGTAGCTT TTTCGAACT AGTATATGTA CATTAGCTGC CTGTGGATTA ACATTTCCCT	180
5	GAATGTAT TTTCGAACT TGAATCTTAA ACTTTTGTG TCTTTATATA AGGTATGCTV	240
	CTTTAAGCA TGAATTTT AAGCACAATA GTTGAAGAC AATCT/CACC TTTTACTTGT	300
10	ATATTACAT GTACTTAAT TTTTATGCA TATTACGTCT TATTATTTAA CCAACCTATT	360
	TTCTTTATC TGGGGCATTT TCGAAGAGC GTTATTTCT TGTATTAATC AAATATTTTT	420
	ATGATGTAT TTTCCTTAT TATTAGKAA TACGKTACYC YAAATATATA TTGTGGSTAT	480
15	TTTCAGAAAT GCATATGCC TCCTTAATTT ATTAGAGGCT AACCTAAAT ATTACTTTTA	540
	CCACTTACTT GAATTTCTG GAATTTAGA ACATTTATTG TTTTATGCAT TTTAATTCTA	600
20	CTTGTACTTT TACTACTGCT AAACATTATT ATTGTTTTAG ACAAGCCAA ATATATNTTG	660
	TTATTACCTT ATCTCCAT TTTTCTGTA TTTTATGCC ACTATGTATG CTCAATTTCC	720
	TTCTATGTA TGAACCTAAT TCACTACTTT TGTTTTTTAA TCTGTGAGG TAGCCTGGCC	780
25	ATTAATCTT TATTTTGGT TTCTGAAAA AATTGTGTTT ATTCTATAT GCATACTTAT	840
	GCAATAGAA TCTAGTTTG AATATTTTT AGTATTTATA AATGTAAAGT CATTWATTKG	900
30	GCTTCTATCA TTCTGTHGA GAATCAATT GTCAGCCCAA TAGTTTTCA TTTTAAATTA	960
	CGAATCTT TCACTTTCT GTTTTAGGA	990

35

(2) INFORMATION FOR SEQ ID NO: 133:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1720 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

45

	GTCTGATAG CGACTTGGT TATCCCTTA AAGTTTACTT CAGCACTAAC ACTAGTGCTT	60
	CCGCTGGAGT TTGCACTTTT CAGCTTTAT ACAGGATTTT CTTTGGACTG GAAGAGTCAA	120
50	GGATATGAG ACTCAAGCT GACATTTATT GTACAACATC AAGGGGAATA GGATACTCAT	180
	CAACTGGGA TTATCTTAT CAAACATGG TCTTCTTGA ATAAGAAAA TACATAGTTG	240
55	GTTATATGG ACTTAAACT GTSTAAATG GATATTTCTA TAAATATTT GCTGCTCTGT	300
	AGATGTGGA AATCTGGA ATATTAGCTT TACTCATCTT GAGCTTTGAG GATGTTCTCT	360
	GTAGCCCGAT GGTTCCTAT TACTAAAAA AGCTGGGTAT TGTAAATCT CATTATATAA	420
60	AATCAGATG AGAAGAAAT TTTCTTTGAT GGTGAGACTG TTGCTTAGT TCAGGAAAT	480